

Force Field Comparison through
Computational Analysis of Capsular
Polysaccharides of *Streptococcus*
pneumoniae Serotypes 19A and F

Marc Gordon

Department of Computer Science

University of Cape Town

In partial fulfilment of the
requirements for the degree of
MSc in Information Technology

February 2014

Plagiarism Declaration

I know the meaning of plagiarism and declare that all of the work in this dissertation, save for that which is properly acknowledged, is my own.

Acknowledgements

A monumental thank you to my supervisors for the effort they have put in and the myriad of interesting reference papers you both have routinely forwarded over the years. Assoc. Prof. Michelle Kuttel, thank you for your encouragement, reassurance, guidance and tireless dedication to the proof-reading of countless drafts. Words cannot express how helpful you have been. This could not have been possible without you. To Assoc. Prof. Neil Ravenscroft, your unbridled positivity, enthusiasm and warm personality are truly infectious, a meeting with you never failed to fill me with a desire to press on. Without you I fear I may have given up long ago.

To my friends and family that have put up with me on this journey I say thank you. To my father who, despite his limited chemistry background, soldiered through proof-reading several iterations of this thesis I say thank you and well done. To my mother and my brother, thank you for lightening the load for me. It made all the difference. To my loving fiancée Emily, thank you for the endless support, encouragement and insistence that I take a break every now and then. I know this hasn't been easy. To the friends who have had to tolerate me staying home to write instead of coming out on so many occasions I say thank you for your patience and tolerance. To Craig Belfour for being in your office to chat for a few minutes before each meeting with my supervisors.

Thank you to Andrew Lewis at the ICTS-HPC for getting me up and running on the cluster and for tolerating the prolonged testing involved in getting the NAMD platform just right. I hope you and your team keep up the great work you do. To Audrey Mbogo and Sonja Berman for putting up with my constant emails and for making the administration of this degree such a breeze. Thank you as well to the University of Cape Town and the Department of Computer Science for the opportunity to complete this thesis.

I would also like to acknowledge the National Research Foundation for the Scarce Skills Scholarship and the University of Cape Town for the various bursaries bestowed upon me.

Thank you.

Abstract

Modern Molecular Dynamics force fields, such as the CHARMM36 and GLYCAM06 carbohydrate force fields, are parametrised to reproduce behaviours for specific molecules under specific conditions in order to be able to predict the behaviour of similar molecular systems, where there is often no experimental data. Coupled with the sheer number available, this makes choosing the appropriate force field a formidable task. For this reason it is important that modern force fields be regularly compared.

Streptococcus pneumoniae is a cause of invasive pneumococcal disease (IPD) such as pneumonia and meningitis in children under five. While there are over 90 pneumococcal serotypes only a handful of these are responsible for disease. Immunisation with the conjugate vaccine PCV7, has markedly decreased invasive pneumococcal disease. Following PCV7 immunisation, incidences of non-vaccine serotypes, especially serotype 19A, have increased. Serotype 19F's capsular polysaccharide differs from 19A's at a single linkage position. Where 19A possesses an α -D-Glcp-(1→3)- α -L-Rhap (G13R), 19F possesses an α -D-Glcp-(1→2)- α -L-Rhap (G12R) linkage. For this reason it was thought that a 19F conjugate would cross protect against 19A. Unfortunately PCV7 vaccination appears to have been largely ineffective against 19A disease.

The lack of conformational information for the G12R and G13R disaccharides provided a good opportunity to compare the CHARMM and GLYCAM force fields. The dynamics of the G12R and G13R disaccharides were investigated under both CHARMM and GLYCAM. While we did identify some discrepancies, overall the force fields were in agreement in predicting a more flexible G12R than the more restricted G13R. While it is possible that these differences account for the lack of 19F to 19A cross protection, further research is required.

Table of Contents

Chapter 1: Introduction.....	1
1.1. Problem Statement.....	5
1.2. Aims.....	5
1.3. Research Questions.....	6
1.4. Approach.....	6
1.5. Thesis Overview.....	7
Chapter 2: Carbohydrates & Conjugate Vaccines.....	9
2.1. Carbohydrate Chemistry and Classification.....	9
2.1.1. Monosaccharides.....	9
2.1.2. Oligosaccharides, Polysaccharides and the Glycosidic Bond.....	12
2.2. Carbohydrates and Immunity.....	14
2.2.1. Capsular Polysaccharides.....	14
2.2.2. Conjugate Vaccines.....	15
2.3. The Rise of Pneumococcal Conjugate Vaccines.....	16
Chapter 3: Molecular Modelling.....	18
3.1. Molecular Dynamics.....	18
3.2. Carbohydrate Parameter Sets.....	19
3.2.1. CHARMM.....	20
3.2.2. GLYCAM.....	21
3.2.3. GROMOS.....	21
3.3. Metadynamics Sampling.....	22
Chapter 4: Streptococcus pneumoniae.....	26
4.1. Pneumococcal Serotypes 19A and 19F.....	26
4.1.1. Pneumococcal Non-vaccine Serotype Replacement.....	28
Chapter 5: Methodology.....	30
5.1. Approach.....	30
5.2. Molecular Designations.....	30
5.3. MD Simulation Parameters.....	32
5.4. Metadynamics Simulations.....	32
5.5. MD Analysis.....	33
5.6. Oligosaccharide Extensions.....	34
5.7. Software and Hardware.....	35
5.8. Streptococcus pneumoniae Analysis Package.....	36
Chapter 6: Results and Discussion.....	38
6.1. Force Field Comparisons.....	39
6.1.1. G12R (19F).....	39
6.1.2. G13R (19A).....	47
6.2. Linkage Comparison.....	54
6.3. Oligosaccharide Extensions.....	57
Chapter 7: Conclusions and Future Work.....	60
Chapter 8: References.....	62
Chapter 9: Appendices.....	70
Appendix A: NAMD Configuration Files.....	70
1. G12R_C Unbiased Simulation Configuration File.....	70
2. G12R_C Biased Simulation Configuration File.....	71
Appendix B: NAMD Colvars Module Configuration Files.....	72
1. G12R_C.....	72

Chapter 1: Introduction

The extreme flexibility of carbohydrates often makes them difficult to characterise experimentally. Experimental techniques employed in this field are nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography. These techniques introduce some potential problems. X-ray crystallography involves obtaining a static structure of the molecule in a crystalline or solid state. Unfortunately, the flexible nature of carbohydrates means they often adopt a number of conformations while in solution, as opposed to remaining in a single static conformation. It is difficult to know whether or not the molecule persists in the crystalline conformation when in solution. On the other hand, while NMR spectroscopy does provide data on the conformation of carbohydrates in solution, it presents an average of all conformations throughout the duration of the NMR investigation^{1,2}. Other investigative techniques such as electron microscopy, neutron diffraction and infrared spectroscopy also suffer from similar shortcomings².

Molecular Dynamics (MD) simulates atomic interactions using a force field, a series of functions based on Newton's equations of motion, and a set of parameters describing molecular behaviours in systems involving hundreds of thousands of atoms. The MD approach allows for analysis of carbohydrates without the need for averages or crystalline structures. Instead the carbohydrate can be observed in a time-lapse style series of frames known as a trajectory. The parameter sets are created through consideration of both quantum mechanics and the above mentioned experimental techniques³. The results of MD investigations are also often analysed in conjunction with NMR or X-ray Crystallographic data^{1,4-6}. In light of the origins of MD parameter sets and the credence given to the results of experimental techniques, MD is better viewed as complementary, rather than wholly alternative to experimental techniques³.

The quality of the force field is a key determinant of the reliability of molecular simulations. Unlike protein force fields that have benefited from research as far back as the 1970s⁷, carbohydrate MD is relatively new. The GRONingen MOlecular Simulation (GROMOS)⁸,

CHARMM⁹ and GLYCAM¹⁰ force fields arguably represent the dominant force fields in current day carbohydrate research. The first CHARMM and GLYCAM carbohydrate parameter sets were only published in 1988¹¹ and 1995¹² respectively. Modern carbohydrate force fields are still missing key parameters and still produce results in disagreement with one another¹³. It is therefore of crucial importance that modern force fields be routinely compared in order to not only guide future force field development but to aid researchers in choosing the appropriate force field for their investigations.

The CHARMM and GLYCAM force fields have similar aspects in their parametrisation philosophies. While the CHARMM carbohydrate force field focused on development of a force field that is compatible with other currently utilised force fields, GLYCAM06 (the newest GLYCAM force field¹⁰), attempted to keep its parameters consistent with past AMBER protein parameters. This would allow for more efficient simulation of heterogeneous systems. In the case of CHARMM, the same approach used in CHARMM force fields concerning other biomolecules was utilised in the carbohydrate force field. Both CHARMM and GLYCAM06 are parametrised primarily for use in aqueous solutions, with CHARMM non-bonded parameters developed using solute-water interactions and all simulations under GLYCAM06 taking place using TIP3P water models (CHARMM also utilises the TIP3P water model as opposed to GROMOS' favouring of the SPC model^{2,14,15}). Both CHARMM and GLYCAM06 are further characterised by a lack of 1-4 non-bonded interaction scaling^{10,16,17}. These common elements help to facilitate comparison.

Invasive pneumococcal disease (IPD) is one of the leading threats to young children today, with nearly 1 million deaths per year in children under 5 years old¹⁸. IPD is the result of infection by the bacteria *Streptococcus pneumoniae* (*S. pneumoniae*). Depending upon the nature of the infection, IPD can manifest in a number of ways, arguably the worst being meningitis, pneumonia and otitis media. IPD is a significant killer of children below the age of 5, with the majority of these deaths occurring in emerging countries¹⁸. In 2000 alone, just over 800 000 children died as a result of IPD, with 95% of those deaths occurring in developing African and Asian countries¹⁸. Pneumococcal attack rates in individuals that are HIV positive are 40 times higher than those that are seronegative¹⁹. *S. pneumoniae* potentially poses a greater risk to the populations of countries with high HIV incidence

rates such as South Africa, than in developed countries where less of the population is HIV positive.

S. pneumoniae is a gram negative bacterium wrapped in a protective Capsular polysaccharide (CPS) coating. Polysaccharides are long chain carbohydrates made up of a number of units termed monosaccharides. Numerous monosaccharides exist each with a number of stereoisomers. The monosaccharides bind together in a variety of formations through linkages called glycosidic linkages. All glycosidic linkages have the dihedral angles (torsion angles defined using four atoms) Φ (H1-C1-O1-Cx) and Ψ (C1-O1-Cx-Hx) (Figure 2.6). Regarded as the major source of carbohydrate flexibility²⁰, these dihedral angles are a frequent subject of carbohydrate conformation investigations involving the analysis of the Φ and Ψ angles in the form of free energy surfaces (FESs) depicted as Ramachandran style plots²¹ (a 2D plot with Φ , Ψ values on the X, Y axes respectively and free energy values depicted using coloured isocontour lines)^{4,5,22,23}. Owing to the need to fully explore the entire FES sampling techniques like Metadynamics²⁴ are used in addition to standard MD approaches. Thus far, over 90 pneumococcal serotypes have been identified and this number continues to grow, with the serotype 6D having been identified as recently as 2009^{18,25}.

Even though only a handful of the 90+ serotypes are responsible for severe IPD, the number of problematic serotypes are higher than in most other diseases combated with vaccines. For this reason, an effective pneumococcal vaccine must contain a high number of serotypes (have a high valency). With a valency of seven, Pfizer's PCV7 (marketed as PrevnarTM/Prevenar^{TM26}) was the first pneumococcal conjugate vaccine to be licensed. It contains the pneumococcal serotypes that, at the time, were responsible for most of the IPD in the United States. While PCV7's efficacy varied greatly from region to region²⁷ it has been very successful, reducing the incidence of IPD significantly, especially with regard to vaccine serotype IPD²⁸.

The post PCV7 era has been characterised by the rise of serogroup 19 as one of the dominant causes of *S. pneumoniae* infections worldwide. While the serogroup consists of serotypes 19A, 19B, 19C and 19F,, serotypes 19A and 19F are the major causes of disease²⁹.

19A: $[\rightarrow 4)\text{-}\beta\text{-D-ManpNAc}\text{-(1}\rightarrow 4)\text{-}\alpha\text{-D-Glcp}\text{-(1}\rightarrow \mathbf{3})\text{-}\alpha\text{-L-Rhap-1-P}\rightarrow]$

19F: $[\rightarrow 4)\text{-}\beta\text{-D-ManpNAc}\text{-(1}\rightarrow 4)\text{-}\alpha\text{-D-Glcp}\text{-(1}\rightarrow \mathbf{2})\text{-}\alpha\text{-L-Rhap-1-P}\rightarrow]$

The CPSs of 19A and 19F consist of trisaccharide repeating units. The only structural difference between these repeating units occurs in the bold portion of the structures above. 19A has a 3-linked glycosidic bond between the glucose and rhamnose residues while 19F possesses a 2-linked glycosidic bond^{30,31}.

The stability of 19F meant it was included in the vaccine over 19A. It was believed at the time that with only a single difference between their CPSs, 19F could provide cross protection for 19A. Cross protection means that a vaccine containing just one serotype might illicit antibodies capable of fending off multiple serotypes. Unfortunately despite some promising research^{32,33}, the post-PCV7 era has been characterised with a rise in non-vaccine serotype prevalence, especially 19A^{34–37}. This rise is now attributed to non-vaccine serotype replacement³⁵. The serotype replacement created the need for newer vaccines, PCV10 (marketed by GlaxoSmithKline (GSK) as Synflorix^{TM26}) and PCV13 (marketed by Pfizer as Prevnar13TM/Prevenar13^{TM26})^{28,34,37,38}. While PCV13 contains a 19A conjugate, PCV10 does not²⁶.

The decision of which conjugates to include in a vaccine ideally occurs on the basis of a comprehensive understanding of the 3D structure of each serotype and how the human immune system responds to each. There is currently little information, both experimental and computational, regarding the dynamics of 19A and 19F CPSs.

1.1. Problem Statement

While protein force fields have in the past been routinely examined^{39–41}, carbohydrate force fields are reviewed less frequently. Further, comparisons of the newer carbohydrate-specific force fields, such as GLYCAM06, in their application to disaccharides, are exceedingly rare. Over the years comparisons between both the current and previous generations of CHARMM and GLYCAM force fields have been undertaken with varied degrees of conformity in the results^{13,42,43}. One must keep in mind that most force fields are parametrised to reproduce specific molecular behaviours in specific environments given a sample of molecules. The aim is to predict the behaviours of similar molecular systems for which there are often little to no experimental data⁴². There are a variety of force fields to choose from making the choice of the optimum force field difficult, especially for inexperienced users. Choosing the optimum force field for the task at hand is however extremely important. This makes ongoing comparisons between these dominant force fields crucial.

Little is known about the favoured conformations of 19A and 19F CPSs. There have been experimental investigations performed but to date^{30,31,44} only a couple of computational investigations have been undertaken^{4,45}. This provides the perfect opportunity to perform a comparison of leading carbohydrate-specific force fields while obtaining valuable data on the conformational dynamics of the 19A and 19F CPSs.

1.2. Aims

The aims of this work are twofold: To compare and contrast the results produced by the CHARMM and GLYCAM force fields. This will add to the knowledge of force field behaviour under a variety of conditions and help future molecular dynamics users to better select the optimum force field for their needs.; To perform a molecular dynamics structural investigation of serotype 19A and 19F in order to better understand the structural relationships of these serotypes.

1.3. Research Questions

1. **How does the choice of force field affect the calculated conformations of the α DGlup- α LRhap disaccharides?** In the past CHARMM and GLYCAM have been compared in disaccharide conformation investigations. The CHARMM parameters were demonstrated to produce dissimilar results to those of GLYCAM¹³. What impact does the utilisation of a different force field have on the modelling of the disaccharide conformations?
2. **What effect does the 19A and 19F trisaccharide structural difference have on the conformation of the α DGlup- α LRhap disaccharide?** The PCV7 conjugate vaccine did not include a 19A serotype conjugate as it was believed that 19A and 19F CPSs were structurally similar enough that an adequate degree of cross protection could be obtained through a single serotype conjugate. The subsequent increase in 19A prevalence has been attributed to a lack of cross protection and subsequent non-vaccine serotype replacement³⁵. It would appear that the serotypes are not structurally similar enough to allow for cross protection, but how dissimilar are they?

1.4. Approach

To facilitate a force field comparison and identify potential changes between G12R and G13R, this study begins with computational analysis of the 19A and 19F α DGlup- α LRhap disaccharides' glycosidic linkages (G13R and G12R respectively). The simulations are performed under both the CHARMM and GLYCAM force fields. As an initial point of investigation and due to time constraints involved in solution simulations with long durations, all simulations are performed in a vacuum. Both standard MD simulations (unbiased) and simulations employing the Metadynamics sampling method (biased) are undertaken. The biased simulations utilise the Φ and Ψ glycosidic torsion angles as collective variables to generate FESs in the form of Ramachandran style plots. The unbiased simulations are used to create scatter plots to better understand which energy

wells are visited and the transition paths between those wells in both disaccharides under both force fields.

Oligosaccharide chains, termed oligosaccharide extensions, matching the residue sequences of 19A and 19F are created. In order to construct the oligosaccharide extensions, in addition to the G12R and G13R simulations the β DManpNAc- α DGlc_p linkage (M14G) common to both serotypes is also simulated. A lack of availability of ManNAc CHARMM parameters means the M14G disaccharide is simulated using only the GLYCAM force field. The minima obtained from the disaccharide FESs are used to construct oligosaccharide extensions consisting of chains of six 19A or 19F trisaccharide repeating units. The lack of viable phosphate parameters at the time for both the CHARMM and GLYCAM force fields means that these extensions are not necessarily true representations of 19A and 19F CPSs. They do however highlight the potential impact the choice of force field, as well as the single G12R and G13R structural difference, could have on CPS conformations.

As a further comparison point between the disaccharides and the force fields, the behaviour of glucose's primary alcohol group is analysed. The conformation of the primary alcohol group is defined through the ω torsion angle. The most favourable conformations throughout all MD (unbiased) and Metadynamics (biased) simulations are identified in both disaccharides under both force fields. The primary alcohol group in unbiased isolated glucose will be similarly analysed to determine what effect, if any, the presence of the rhamnose residue has on favourable ω angles.

1.5. Thesis Overview

Chapter 2 introduces the biochemistry and methods of classification of carbohydrates. The relationships between carbohydrates and immunity as well as the history of conjugate vaccine development are also discussed.

Chapter 3 provides background information on the various aspects of the molecular dynamics methodologies employed in the study. General molecular dynamics principles

are discussed, as well as the basics of the Metadynamics sampling method. Brief development histories of the CHARMM and GLYCAM force fields are also included.

Chapter 4 introduces the molecular structure of serotype 19A and 19F CPS repeating units and how those structures fit into the problem of non-vaccine serotype replacement. Past research into these serotypes is also considered.

Chapter 5 details the methodology of the project, addressing hardware and software utilised, disaccharide naming designations, methods for oligosaccharide extensions, molecular dynamics simulation settings of importance and in-house software applications developed and utilised throughout the duration of this study.

Chapter 6 presents the results and discusses their implications. FESs produced under CHARMM and GLYCAM forcefields for both G12R and G13R are analysed and compared. The structure of the oligosaccharide extensions are compared and contrasted by force field and glycosidic linkage. The behaviour of the primary alcohol group will also be investigated.

Chapter 7 reviews the findings in terms of suggested future work to be performed. The need for new force field parameters and the inclusion of additional force fields for future comparisons are identified amongst other avenues of potential interest.

Chapter 2: Carbohydrates & Conjugate Vaccines

2.1. Carbohydrate Chemistry and Classification

Carbohydrates are so named because many of them can be described by the formula $(\text{CH}_2\text{O})_n$ and are thus hydrates of carbon. This simplistic view is not strictly accurate as many carbohydrates contain other components such as amine, sulphate and phosphate groups.

Arguably the most ubiquitous of all biomolecules, carbohydrates are involved in a large number of organic systems. Some bacteria coat themselves in carbohydrates in an effort to evade the human immune system. So effective is this strategy that humanity has had to embrace a regime of childhood vaccinations in an attempt to control many of these carbohydrate shielded pathogens⁴⁶. Through photosynthesis plants store energy in the form of starch or cellulose. Considering that the starch contained in foods such as rice, wheat and potatoes forms the basis of most humans' diets, the central role of carbohydrates is self evident. Through oxidative metabolism, both plants and animals make use of carbohydrates as their primary energy source. Carbohydrates are a crucial component of both DNA and RNA in the form of deoxyribose and ribose, The carbohydrate chitin is a key structural component making up the exoskeletons of arthropods, beaks and gladii of cephalopods, cell walls of fungi and radulae of molluscs. In addition to these structural roles, carbohydrates are also found as components in many glycolipids and glycoproteins involved in a wide range of processes from molecular recognition and cell signalling, to protein stabilisation and cryoprotection. ^{22,47–49}

2.1.1. Monosaccharides

Monosaccharides are both the simplest form of carbohydrate and the basic building block for the more complex carbohydrates. As such, monosaccharides are also important for carbohydrate classification. The number of monosaccharides a carbohydrate consists of is

used to categorise. A carbohydrate with two monosaccharides is known as a disaccharide, with three the term trisaccharide is used, between two and twenty the term oligosaccharide is used and beyond twenty the term polysaccharide is used. Monosaccharides usually conform to the $(\text{CH}_2\text{O})_n$ formula with n being no smaller than three.

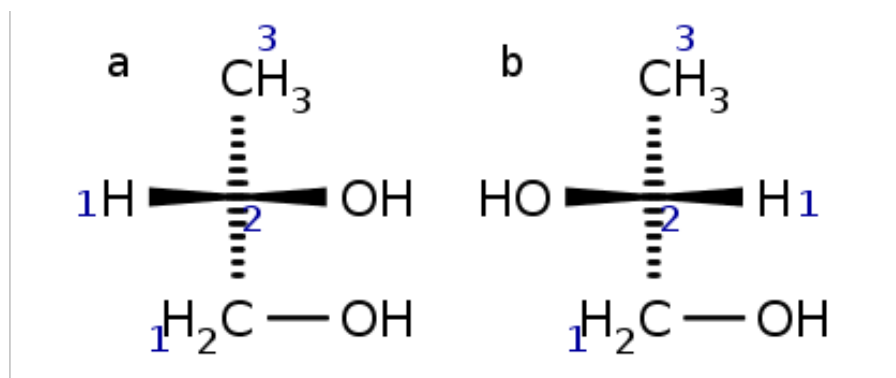


Figure 2.1. *D*-glyceraldehyde (a) and *L*-glyceraldehyde (b)

The distinction between D-L stereoisomers is demonstrated by monosaccharide glyceraldehyde. In Figure 2.1 we see glyceraldehyde's D-L stereoisomers, D-glyceraldehyde and L-glyceraldehyde. D-L stereoisomers are determined by the orientation of the chiral carbon, in this case C2. When the hydroxyl group (OH) of that carbon is pointing to the right, that is indicative of the D stereoisomer, while a left-pointing hydroxyl characterises an L stereoisomer. Where more than one chiral carbon is present the chiral carbon furthest from the carbonyl group is used. A handful of L-sugars are found in nature but the D enantiomer is dominant with most biochemical processes configured for the use of D-sugars. ⁴⁷

Carbon chains of five or more atoms have the potential to form rings. Rings containing less than five atoms are sterically strained. Rings containing five (furanose) and six (pyranose) atoms are common. Although rings of more than 6 atoms do occur in nature, for the most part they serve only minor roles. ⁴⁷

Even within the rings themselves there is yet more conformational variation. An example of this variation can be seen in Figure 2.2. The asymmetrical centre introduced at carbon 1 results in stereoisomers known as α and β anomers. In the α anomer, the hydroxyl group

on carbon 1 exists below the ring while in the β anomer the hydroxyl group occurs above the ring. In a process called mutarotation, monosaccharides are able to convert between their α and β anomers using an intermediary open chain stage. This results in purified anomers approaching an equilibrium mixture in an aqueous solution.

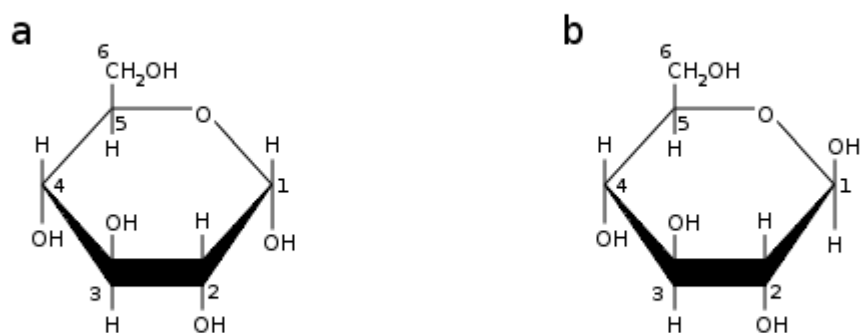


Figure 2.2. The α (a) and β (b) anomers of glucopyranose

Pyranose rings are not planar, but have a defined pucker, the most common of which is the 4C_1 chair (Figure 2.3). We do not consider other ring puckers in this work, such as the inverted chair or boat, as they are higher in energy⁵⁰.

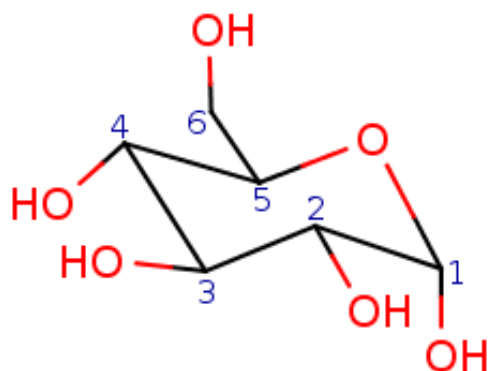


Figure 2.3. α -D-glucopyranose in the 4C_1 formation

The behaviour of the dihedral angle of the pyranose's 6-hydroxymethyl, also known as the primary alcohol group, under CHARMM derivative force fields has been the subject of investigation in the past²⁰. In the case of the primary alcohol the dihedral is defined by either O5-C5-C6-O6 (ω) (Figure 2.6) or C4-C5-C6-O6 (ω_2). There are typically three favourable staggered conformations for a primary alcohol group's dihedral angles. These three conformations are termed trans, gauche⁺ (g⁺) and gauche⁻ (g⁻). The ω dihedral angle is 180° in trans, 60° in g⁺ and -60° in g⁻. Figure 2.4 below shows the relative positioning of the atoms involved in the ω dihedral with C₄ used for positional reference.²⁰

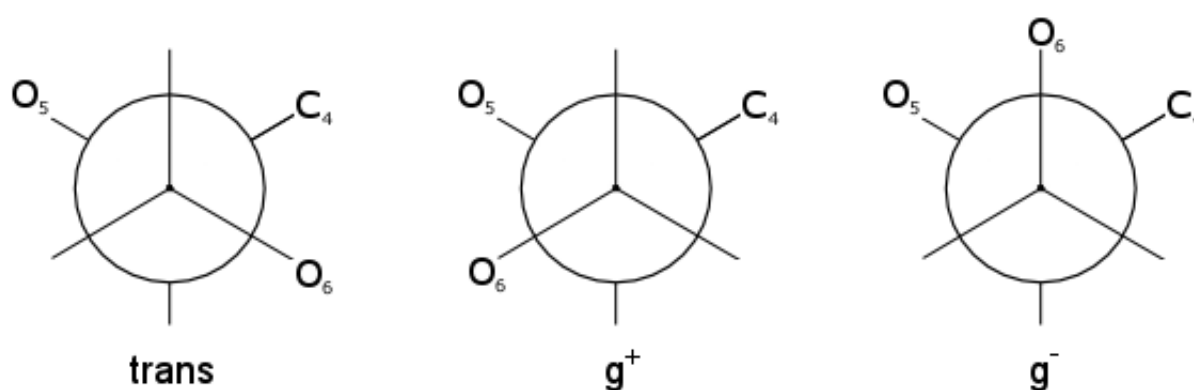


Figure 2.4. *Trans, g⁺ and g⁻ primary alcohol conformations*

2.1.2. Oligosaccharides, Polysaccharides and the Glycosidic Bond

As mentioned previously, a saccharide chain consisting of two to twenty monomers is regarded as an oligosaccharide. Beyond twenty monomers the chain is classified as a polysaccharide. An oligosaccharide consisting of two monomers is known as a disaccharide while one with three monomers is termed a trisaccharide.

Monosaccharides are capable of forming bonds through the elimination of water in a reaction involving the monosaccharide's anomeric hydroxyl and the hydroxyl group of a second compound. The resulting ester bond is known as a glycosidic bond. This second compound could be any number of hydroxyl containing compounds, including another monosaccharide.⁴⁷

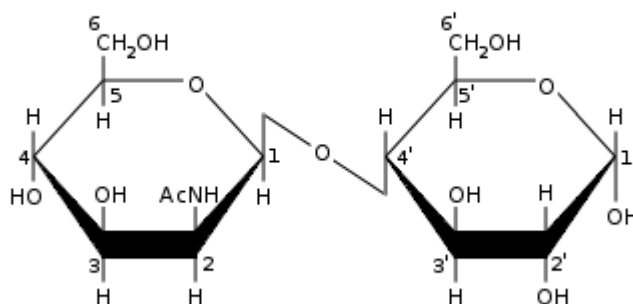


Figure 2.5. *Disaccharide exhibiting a glycosidic bond*

When two monosaccharides are bound together through a glycosidic linkage a disaccharide is formed (Figure 2.5). With rigid ring structures the flexibility of carbohydrates comes mainly from these glycosidic linkages so although disaccharides are relatively simple, they contain many of the same aspects influencing conformation and flexibility in more complex saccharide chains²⁰.

Depending upon the monosaccharides involved, there is potential for glycosidic bonds to form between any of the carbons (Figure 2.5 involves a 1-4 bond). Provided the second monomer's anomeric carbon does not take part in glycosidic bond formation new monomers can be added. In this manner the glycosidic linkage forms the backbone and branches of saccharide chains.

All glycosidic linkages possess torsion angles. Torsion angles of consistent importance in carbohydrate investigations are ϕ and ψ (Table 2.1 and Figure 2.6).

Torsion Angle	Constituent Atoms
ϕ	H1-C1-O1-Cx
ψ	C1-O1-Cx-Hx

Table 2.1. *The atoms making up the ϕ and ψ glycosidic torsion angles*

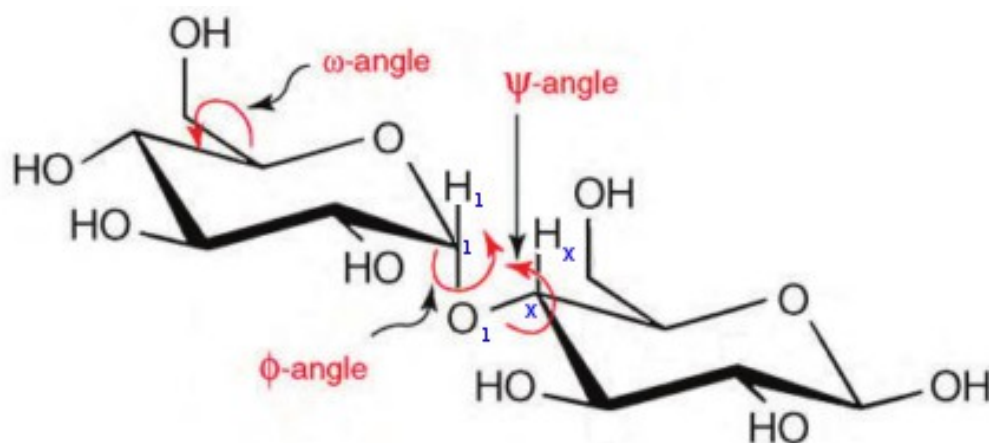


Figure 2.6. ϕ , ψ and ω torsion/dihedral angles in maltose (picture adapted from Fadda and Woods³)

While pyranose rings possess a certain amount of flexibility, the influence this flexibility has over carbohydrate conformation is small. Of far greater influence are the ϕ and ψ glycosidic torsion angles, which define the orientation of a glycosidic bond which determines the conformation of saccharide chains as a whole. It is for this reason that investigations into carbohydrate conformations tend to focus on the ϕ and ψ glycosidic torsion angles^{4,5,22,23} and why analysis of component disaccharides is often regarded as an ideal initial phase in the investigation of the dynamics of more complex carbohydrates⁶.

2.2. Carbohydrates and Immunity

2.2.1. Capsular Polysaccharides

Carbohydrates play a key role in the evasion of the human immune system by some bacterial pathogens. These bacteria form capsules consisting of polysaccharides that cover the surface of the bacteria. Depending upon the source bacteria, these CPSs typically consist of repeating units of between one and six sugar residues⁵¹.

The sugary coating forms a barrier that has been demonstrated to provide cryoprotection and protection from dessication^{52,53}. Some adopt structures similar to those found in the human body. With a polysaccharide coating of similar sequence to those found in humans, the immune system is less likely to recognise the bacteria as foreign⁵¹. The capsule also prevents opsonisation, the binding of antibodies to a pathogen to mark it for disposal.

Without opsonisation the immune system's complement pathways (a set of proteins that function together to aid antibodies and phagocytes in the clearing of pathogens) cannot be initiated and phagocytosis by cells such as macrophages is impeded. There are various mechanisms by which this occurs: Components that would usually allow for alternative complement pathway activation are instead covered up by the capsule; Sialic acid contained within some capsules causes preferential binding for certain serum proteins that inhibit the alternative complement pathway; Some capsules decrease binding affinity for Factor B thereby stopping the complement cascade midway⁵⁴. Some bacterial carbohydrate capsules are so effective that mankind has had to resort to vaccination to control those pathogens.

2.2.2. Conjugate Vaccines

In its simplest sense, production of an effective and safe vaccine involves separation of the disease causing components of a germ, from the components that illicit an immune response. Once this is achieved, the components that illicit an immune response can be administered to induce immunity whilst avoiding the deleterious effects of the germ.

It has already been well established that a comprehensive immune response against bacterial CPSs confers immunity against the associated disease^{55,56}. In vulnerable groups, such as the very young and the elderly, the immune system's response to most CPSs progresses in a manner that results in no long term immunological memory. Meaning that usually when someone below the age of 2 or above the age of 65 is vaccinated using a pure polysaccharide vaccine, the immunity lasts only a short period of time⁵⁴. The lack of lasting protection for vulnerable age groups renders pure polysaccharide vaccines unable to effectively combat bacteria such as *S. pneumoniae* alone.

A failure to generate immunological memory upon exposure to a CPS can be sidestepped by attaching a protein (known as a carrier protein) to the CPS⁵⁷. The exact mechanisms behind this approach are not fully known. This same principle however is used in modern conjugate vaccines. The poor CPS immune response in vulnerable groups is avoided by making use of the immune responses typically associated with proteins.

2.3. The Rise of Pneumococcal Conjugate Vaccines

S. pneumoniae was first isolated in 1880 by both Louis Pasteur (Paris) and George Sternberg (New Orleans) independent of one another. In 1886 pneumococcus was identified as the primary cause of community-acquired bacterial pneumonia. A year later it was unknowingly observed that administering killed pneumococci to rabbits conferred immunity to *S. pneumoniae*^{19,58}.

Despite these early observations, it took a threat to the viability of South Africa's gold industry three decades later to finally galvanise researchers into serious investigation of vaccination against *S. pneumoniae*. Pneumonia attacks in South African mining compounds were approaching 100 cases per 1000 people with a 25% fatality rate. In 1911 testing of a heat-killed whole cell pneumococcal vaccine on miners living in the mine compounds began⁵⁹. The efficacy reported in this trial and the trials that followed remain a point of contention⁵⁹.

In the mid to late 1920s the capsules surrounding pneumococcal bacteria were identified as carbohydrates and their immunogenicity was confirmed. Research began into pure polysaccharide vaccines, which proved to be rather efficacious, leading to licensing of both an adult and a paediatric hexavalent vaccine in the 1940s. These licences lapsed soon after due to lack of use following introduction of antibiotic treatments rendering *S. pneumoniae* a controlled issue in the eyes of the public^{59,60}. The emergence of antibiotic resistant pneumococci and the identification of key problems such as high mortality rates in those over 50, renewed interest in prophylaxis. A 14-valent polysaccharide vaccine was licensed in the United States in 1978 with a 23-valent vaccine following soon after in 1983^{59,61,62}. Merck's Pneumovax 23 is still in use today containing serotypes 1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19F, 19A, 20, 22F, 23F and 33F. The serotypes in the Pneumovax 23 vaccine represent 85-90% of the serotypes involved in IPDs^{19,63}.

The previously mentioned shortcomings of polysaccharide vaccines in infants and the elderly led to the need for the current generation of pneumococcal conjugate vaccines. The first conjugate vaccine produced was Wyeth's Prevnar 7 (PCV7) in 2000. Until recently, PCV7 has been the stalwart conjugate vaccine licensed for use against *S. pneumoniae*. Containing the majority of disease causing serotypes in the United States at the time (4, 6B, 9V, 14, 18C, 19F, and 23F), this vaccine's efficacy varied widely with an estimated greater than 80% efficacy in North America and Oceania, 40% in Asia and 60% in Africa and Latin America^{27,61}. This varied efficacy, along with the occurrence of non-vaccine serotype replacement³⁵ led to the introduction of the most recent vaccines for *S. pneumoniae*, PCV10 and PCV13^{28,37,64}.

Looking to the future, Merck has already begun trials on a 15 valent pneumococcal conjugate vaccine⁶⁵.

Chapter 3: Molecular Modelling

3.1. Molecular Dynamics

Molecular Dynamics (MD) involves the use of computers to simulate the movement of atoms or groups of atoms in a system. MD relies on Newtonian equations of motion. Of particular importance is Newton's second equation of motion:

$$F_{\alpha} = m_{\alpha} a_{\alpha} \quad (1)$$

In equation 1, F is the force exerted upon a particular particle (α) while m and a are the respective mass and acceleration of the particle. F_{α} , expressed as a gradient of potential energy produces the following equation:

$$F_{\alpha} = -\nabla V \quad (2)$$

In equation 2, V is the system's total potential energy. Molecular modelling typically uses a pairwise force expression, where the system's total potential comprises of all covalent bonds, angled bonds (pairs of covalent bonds sharing a common atom), dihedral bonds (atoms separated by precisely three covalent bonds leading to the formation of the torsion angle ϕ) and non-bonded interactions (electrostatic interactions and van der Waal's forces). The heart of any modern MD application is the force field. The force field is a set of parameters that collectively describe the forces experienced by all atoms and the potential energy of the system. If equations 1 and 2 above are combined the result is the following:

$$-\frac{dV}{dr_{\alpha}} = m \frac{d^2 r_{\alpha}}{dt_{\alpha}^2} \quad (3)$$

In equation 3, acceleration has been replaced with atom α 's position (r_α) over time (t). Force experienced by particle α relates to potential energy changes when in different positions. As seen in equation 1, through knowledge of total forces and atomic masses for a particular atom, acceleration of that atom can be determined. The velocity of any particular atom is dependent upon the forces exerted upon that atom by the other atoms within the system. With these calculations, MD applications are capable of determining the positions and acceleration of all atoms systemwide over a series of incremental time steps. This time series of atomic positions is known as a trajectory. A system's trajectory can be obtained by solving differential equation 3. The wealth of information obtainable from a system's trajectory makes it one of the most important outputs of an MD simulation.

Use of the force field parameters to calculate the potential is computationally intensive. Shorter time steps mean a higher degree of accuracy but greater computational requirements. Conversely longer time steps mean less accuracy and lower computational requirements.

Protein and DNA force fields are well established and have benefited from a long period of development. Carbohydrate force fields however have not yet reached the same level of refinement and sophistication. Even some of the dominant carbohydrate force fields have only recently added parameters for complex side groups and also some basic monosaccharide units. CHARMM for example recently added sulphates and phosphates⁶⁶ and did not contain parameters for the rhamnose residue until late 2011.

3.2. Carbohydrate Parameter Sets

Section 3.1 above presents examples forming the basis of the functions involved in MD. The form of the functions utilised differs from one force field to another. The combination of the form of each of those functions, and the accompanying parameter sets describing each atom type, define each force field. The development of these parameter sets are most often an arduous task requiring meticulous attention to detail. While in the past experimental data alone was sufficient for parametrisation, today a combination of quantum mechanical and experimental data are utilised³. The dominant force field

parameter sets in use today all have a rich history, an understanding of that history can help one to choose the correct force field for the task and to put the results of the MD investigation into context.

3.2.1. CHARMM

Maintained currently by Alex MacKerell Jr.'s team from the University of Maryland, the CHARMM series of all-atom force fields were originally developed with the modelling of proteins and nucleic acids in mind⁹. Over time parameter sets for lipids, DNA, RNA, carbohydrates and a handful of other extensions have been added⁹. Although the force field and its accompanying parameter sets were primarily produced for use with the CHARMM MD application, many other MD programs have been developed or extended to utilise them.

The CHARMM carbohydrate force field is the most recent addition to the CHARMM force field series. Parameter set development began with the first CHARMM monosaccharide parameters published in 1988¹¹. In 1993 Grootenhuis *et al.* published a parameter set entitled "CHEAT". CHEAT was CHARMM-based and involved mimicry of a molecule in an aqueous solution through an isolated simulation of that molecule^{67,68}. One of the earliest true carbohydrate parameter sets for CHARMM was presented in 1996. The result of stringing together *ab initio* calculations in small molecular systems such as methanol interacting with water (dubbed "molecular fragments"), the parameter set extended the CHARMM22 force field⁶⁹.

Further extensions served to bring the evolving force field into closer alignment with experimental observations. For example Kuttel *et al.*'s Carbohydrate Solution Force Field (CSFF) addressed a CHARMM derivative's tendency to favour the trans conformation of rotameric distributions and frequencies in solution²⁰.

The CHARMM force field's carbohydrate parameter set is continuously being extended. Some recent extensions include additional parameters for hexopyranose monosaccharides and their glycosidic linkages^{16,70} and more recently the addition of phosphate and sulfate parameters⁶⁶.

3.2.2. GLYCAM

GLYCAM is a series of carbohydrate force fields produced and maintained by Robert Woods' research group at the University of Georgia⁷¹. GLYCAM was originally designed as a parameter set dependant on the MD application AMBER¹⁰. In the early 1990s, to address the lack of carbohydrate parameters in AMBER, a series of carbohydrate parameter sets were produced using the AMBER force field as a base^{68,72,73}. Along with these other parameter sets, the Woods team presented their own parameter set called GLYCAM_93 for use with the AMBER force field¹².

The most recent leap forward is the release of GLYCAM06. Unlike its predecessors, GLYCAM06 is independent of AMBER. The testing and training involved around 100 hydrocarbon molecules from a variety of side group categories including alcohols, amides and ethers along with simple ring structures. If at any point during a torsion rotation an internal hydrogen bond could be formed, two energy curves would be analysed, with one curve permitting the hydrogen bond and another disallowing it¹⁰.

3.2.3. GROMOS

The first major GRONingen MOlecular Simulation (GROMOS)⁸ force field was GROMOS87⁷⁴. This was followed nearly a decade later by GROMOS96⁸.

The 53A6 and 45A4 force field parameter sets were produced in 2004 and 2005 respectively^{14,15,75}. While 53A6 was a general force field and 45A4 was built for use with carbohydrates, the 45A4 parameter set was incorporated into 53A6 making them near identical in pure carbohydrate simulations². These force field parameters have a strong history, being utilised in numerous investigations^{6,76–79}.

The latest development in the GROMOS carbohydrate force fields is designated 56A_{CARBO}. Designed for use for hexopyranose based systems, 56A_{CARBO} is a reoptimised version of the general 53A6 force field. The 56A_{CARBO} parameter set addresses key flaws identified in the 53A6 parameter set².

3.3. Metadynamics Sampling

To be useful, MD simulations should visit as many energetically relevant structural conformations as possible. MD relies on deterministic systems emulating natural forces to bring about conformational change. For this reason MD simulations have limitations. Relying on natural fluctuations in systemic forces leaves the user with little control over how quickly and whether or not a system will adopt relevant conformations.

Relying upon natural forces it may take the system a long time to visit the relevant conformations, making simulations needlessly time consuming. Another possibility is that energy wells on the FES may be separated by high energy barriers that are difficult, if not impossible, to overcome through natural forces. This means that crucial areas of the FES could remain inadequately sampled or entirely unvisited in a standard MD simulation.

To address these issues, certain MD approaches involve the application of external forces to a system. This allows the user to simulate situations that might be rare or even entirely impossible in nature. These artificial situations allow users to produce information that would not be obtainable through physical experimentation. Examples include Umbrella⁸⁰, Adaptive Biasing Force⁸¹ and Metadynamics sampling techniques²⁴.

First introduced by Laio and Parrinello in 2002²⁴, Metadynamics is an enhanced sampling technique usually employed in MD. Metadynamics produces a representation of the FES of a system through the allowance of rigidly defined degrees of freedom such as bond lengths and torsion angles.

The principle behind Metadynamics can be described using the analogy of a blind man trapped in an empty swimming pool with a bucket containing an infinite supply of sand. Unable to see or climb, he cannot escape the deep end of the pool. The man begins pouring sand from the bucket into the bottom of the pool to fill the well of the deep end. When it is sufficiently filled, he is able to walk along the sand and enter the shallow end's well. Over time he again fills the shallow well with sand. Continuing in this manner the blind man eventually fills the entire pool and is able to escape. If he were able to recall

where every bucket of sand were dropped, he would have an opposite image of the pool's topography and would therefore be able to accurately reconstruct the original topography⁸².

With Metadynamics, like the sand, a history-dependant biasing force acts to discourage the system from returning to previously visited conformations. While carbohydrates are flexible, and usually occupy a variety of conformations, they are often separated by significant energy barriers. This makes Metadynamics an invaluable carbohydrate conformational analysis tool when analysing Φ and Ψ glycosidic torsion angles. The Kuttel and Naidoo study⁸³, for example, utilised a biasing force similar to the Metadynamics approach in order to overcome significant energy barriers in the maltose FES. While it is possible that the lower energy barrier could have been overcome given enough time in an unbiased simulation, it is unlikely that the larger barrier, representing an energy difference of over 10kcal/mol *in vacuo*, would have been overcome.

The defined degrees of freedom in Metadynamics are referred to as collective variables (CVs). The choice of appropriate CVs is essential and the nature of those CVs depends largely upon what is being studied. Investigations of carbohydrates involving Metadynamics have utilised numerous CVs including ω primary alcohol dihedral angles⁸⁴ and ring pucker coordinates^{50,85–87}. All CVs have three fundamental requirements: First, CVs must collectively form an accurate description of the conformational transition of interest. In other words they must collectively describe the system's initial state, final state and all relevant intermediary states; Second, the CVs must encompass all variables that could not otherwise be sampled within the simulation's duration. These otherwise slower changing variables need to be fully explored before the simulation's end while the faster changing variables can simply react to the bias-influenced changes in the slower variables; Finally the number of CVs should be kept low. If too many CVs are defined it means the space to be explored in order to produce an FES has many dimensions. This will quickly result in impractical processing times.^{88–90}

The bias potential can be calculated from the following equation:

$$V_G(S, t) = \int_0^t dt' \omega \exp \left(\sum_{i=1}^d \frac{(S_i(R) - S_i(R(t')))^2}{2\sigma_i^2} \right) \quad (4)$$

Equation 4 allows one to calculate the bias potential (V_G) for the set of CVs (S) at time (t). The bold portion concerns the CVs of the system. It originates from the equation 5 involving a set of CVs (S), defined as a number of functions (d) of a set of coordinates (R).

$$S(R) = (S_1(R), \dots, S_d(R)) \quad (5)$$

The biasing force is the sum of deposited Gaussian functions (a type of function characterised graphically as a symmetric bell-curve). The characteristics of the Gaussian are considered in equation 4. While σ_i is the width of the Gaussian of the i th CV, ω is the energy rate expressed as follows:

$$\omega = \frac{W}{\tau_G} \quad (6)$$

where W is the Gaussian height and τ_G is the duration between deposition of Gaussians. Metadynamics algorithms allow for setting of the duration between the addition of Gaussian functions, Gaussian height and Gaussian width in order to strike a balance between accuracy and computation time. This allows the Metadynamics approach to be highly flexible. By changing the characteristics of the Gaussians, the algorithm can be made to produce anything from a quick and rough estimate of the FES, to a high duration simulation resulting in an FES with high granularity.

The Metadynamics sampling technique provides an adaptable and reliable approach ensuring that every portion of the carbohydrate's FES is adequately explored. With the ability to identify energy wells that would otherwise have gone undiscovered, it represents a valuable addition to the MD arsenal.

Chapter 4: Streptococcus pneumoniae

S. pneumoniae, a gram-positive encapsulated bacteria, is one of the main causes of morbidity and mortality with an estimated 1.6 million deaths per annum^{27,91}. Of those 1.6 million deaths somewhere between 700 000 and 1 million occur in children under 5 years of age²⁷. IPD includes *otitis media*, meningitis, bacteraemia and pneumonia^{92,93}. Looking at these brief figures it is clear that *S. pneumoniae* represents a significant threat to human life.

To date over 90 pneumococcal serotypes have been identified with new variants still being discovered in the wild^{25,27,91,94–97}. Despite this large number of serotypes only a fraction are responsible for severe IPDs. Of particular concern are serotypes that are highly prevalent, persist for a long duration, cause IPD and are antibiotic resistant. Against many of these serotypes humanity's major defence comes from conjugate vaccines administered early in life.

4.1. Pneumococcal Serotypes 19A and 19F

Pneumococcal serogroup 19 contains serotypes involved in the development of severe IPDs. Often antibiotic resistant and carried for long durations, members of serogroup 19 can be difficult to treat^{36,95,98,99}. While the serogroup contains serotypes 19A, 19B, 19C and 19F⁹⁶, serotypes 19F and 19A are of particular interest. In the wake of the roll out of PCV7, these two serotypes, in particular serotype 19A, rose to become two of the most common causes of IPD^{35,100}.

Serotype	Structure
19F	→4)-β-D-ManpNAc-(1→4)-α-D-Glcp-(1→2)-α-L-Rhap-1-P→
19A	→4)-β-D-ManpNAc-(1→4)-α-D-Glcp-(1→3)-α-L-Rhap-1-P→

Table 4.1. *S. pneumoniae* serotypes 19A and 19F trisaccharide repeating units

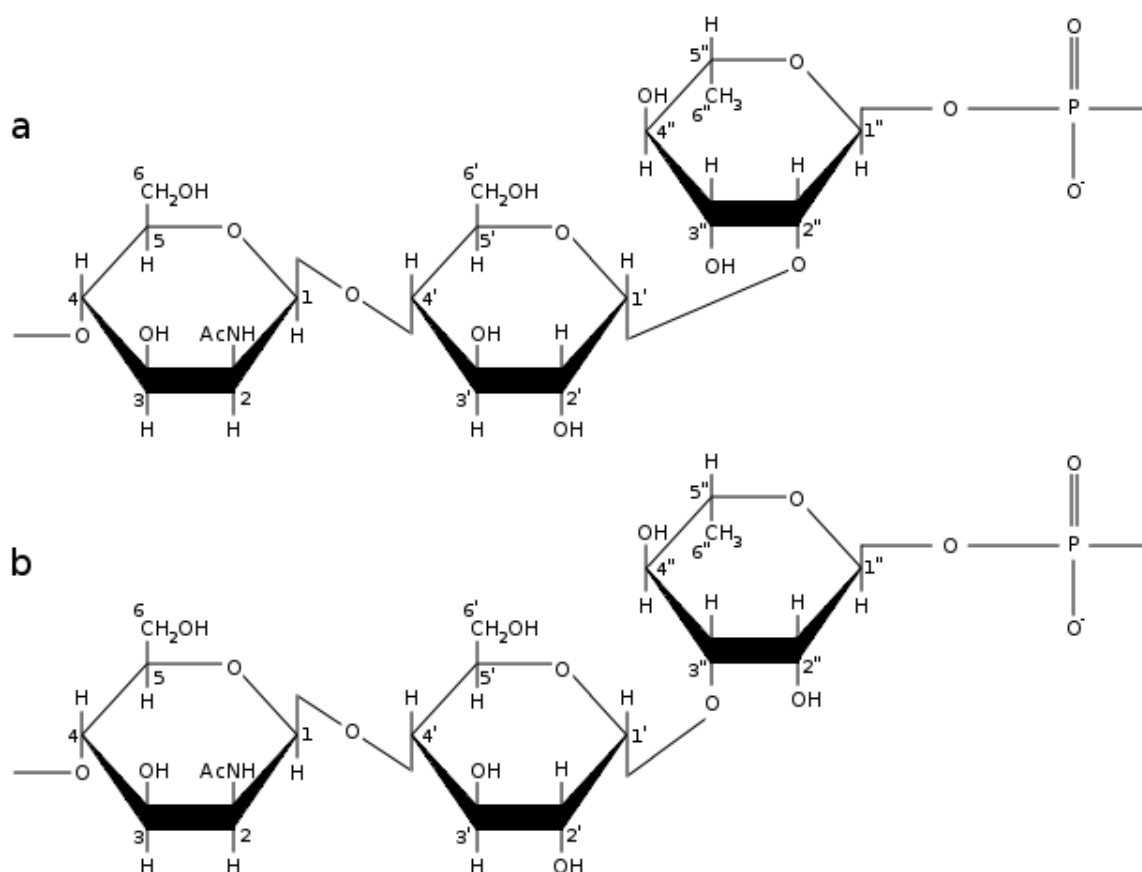


Figure 4.1. 19F (a) and 19A (b) *S. pneumoniae* capsular polysaccharide repeating units

19A and 19F CPS repeating units are structurally similar (Table 5.1 and Figure 5.2) with only a single variation in the glycosidic bond between the glucose and rhamnose residues. In the case of 19A this bond is a 1-3 configuration while in 19F it is substituted with a 1-2 bond^{30,31,101,102}.

There have been a couple of investigations into serotypes 19A and 19F involving use of computational methods. The first such investigation was conducted in 1992⁴. This work utilised the experimental method H-NMR and the now largely outdated computational approach of simple hard sphere (HSEA)¹⁰³ and the MM2 force field¹⁰⁴. The disaccharides making up the 19A and 19F trisaccharide repeating units were investigated (Table 4.2). Using HSEA, the study produced FESs for each of the three disaccharides using ϕ and ψ dihedral angles. MM2 analysis followed with a search around the MM2 identified minima in 1° intervals for the true minima. This approach allowed for the identification of additional global and local minima for each disaccharide.

Serotype	Component Disaccharides
19A	β -D-Man α NAc-(1 \rightarrow 4)- α -D-Glcp
	α -D-Glcp-(1 \rightarrow 3)- α -L-Rhap
19F	β -D-Man α NAc-(1 \rightarrow 4)- α -D-Glcp
	α -D-Glcp-(1 \rightarrow 2)- α -L-Rhap

Table 4.2. 19A and 19F component disaccharides as investigated by Ciuffreda *et al.*⁴

A second MD investigation into serogroup 19 involved a synthesised variation on the serotype 19F trisaccharide repeating unit⁴⁵. This synthesised variation, while maintaining near conformational equivalence to the naturally occurring 19F trisaccharide, exhibited higher stability making it potentially useful for inclusion in a vaccine. As with the previous study, this work did not utilise current dominant carbohydrate force fields such as CHARMM, GLYCAM and GROMOS, instead utilising Hyperchem's MM⁺ force field, based upon the MM2 force field¹⁰⁵ used in the Ciuffreda *et al* study⁴.

Both studies presented rigid disaccharides consisting of glycosidic linkages with limited ϕ and ψ ranges^{4,45}. Only single wells of low energy were observed with Ciuffreda *et al.* observing some limited additional points of low energy without distinct wells⁴.

4.1.1. *Pneumococcal Non-vaccine Serotype Replacement*

PCV7, while undeniably effective, contained only 7 serotypes (4, 6B, 9V, 14, 18C, 19F, and 23F) of the over 90 known pneumococcal serotypes. Prior to the introduction of PCV7, pneumococcal serotypes 19F and 19A were common. Both were also readily associated with IPDs. After the widescale introduction of PCV7 19A rose in prevalence until it became the dominant *S. pneumoniae* serotype in many areas of the world. This is today identified as an example of non-vaccine serotype replacement³⁵.

The change in pneumococcal serotype prevalence figures in the post PCV7 era made it apparent that the vaccine's coverage was inadequate. Pfizer introduced PCV13 as a replacement for PCV7 while GSK introduced its own pneumococcal vaccine, PCV10, all with differing component serotypes¹⁰⁶ (Table 4.3).

Vaccine	Serotypes
PCV7	4, 6B, 9V, 14, 18C, 19F, 23F ¹⁰⁶
PCV10	PCV7 + 1, 5, 7F ¹⁰⁶
PCV13	PCV10 + 3, 6A, 19A ¹⁰⁶

Table 4.3. *Component serotypes of pneumococcal conjugate vaccines*

Even before the introduction of PCV7, whether or not structural similarities between the CPSs of serotypes 19A and 19F might allow one to cross protect for the other was already a contentious issue¹⁰⁷. The advantage of cross protection is that it reduces the cost of the vaccine without compromising its protective capabilities. Unfortunately the 19F conjugate in PCV7 did not provide adequate cross protection for 19A pneumococcal serotypes³³. Today the issue of cross protection between serotypes 19A and 19F remains an issue of concern as PCV13 contains a 19A conjugate while PCV10 does not. Initial investigations have suggested that PCV13 induces immunity to serotype 19A^{108–110}. GSK's PCV10 however utilises a different method for joining the protein and carbohydrate components of the conjugate to those seen in Pfizer's PCV7 and PCV13. It has been argued that GSK's approach could influence cross protection between serotypes 19A and 19F¹¹¹. There have also been numerous cost analysis investigations into the viability of both PCV10 and PCV13 in numerous countries with a variety of conclusions being drawn^{102,106,112}.

There is a lack of investigation into PCV10's ability to provide cross protection for 19A as well as persistent questions hanging over the issue of serotype 19A and 19F cross protection. Coupled with the lack of any MD investigations into these serotypes involving modern dominant carbohydrate force fields such as CHARMM and GLYCAM, the need for further research is readily apparent.

Chapter 5: Methodology

5.1. Approach

This study made use of the NAMD MD application^{113,114}, along with the colvars Metadynamics sampling module, in the investigation of the *Streptococcus pneumoniae* serotype 19A and 19F CPS repeating units. Of importance was a comparison of their behaviour under both the CHARMM and GLYCAM force fields.

In addition to biased Metadynamics simulations, unbiased simulations were also performed on the G12R and G13R disaccharides as well as glucose under both CHARMM and GLYCAM. This facilitated glycosidic linkage comparisons and force field comparisons. These unbiased simulations were performed to observe molecular behaviours in absence of the Metadynamics biasing forces, reveal potential traversal paths between energy wells and other such behaviours not easily observed in biased simulations. In particular, the glucose residue's primary alcohol group was thought to be involved in interresidue bonding patterns and so became an area of interest. Glucose was simulated under both force fields to determine whether or not the behaviour of the primary alcohol group was influenced by the presence of the rhamnose residue.

Oligosaccharide extensions were created for each of the glycosidic linkages under each of the force fields. These extensions allowed for observation of the potential impact of not only the serotype linkage differences but also the relative shifts of the minima across the CHARMM and GLYCAM force fields.

5.2. Molecular Designations

Due to a lack of adequate parameters in both the GLYCAM and CHARMM forcefields, it was decided that the isolation of the disaccharide pair exhibiting the single difference between the 19A and 19F CPSs would be the main focus of the investigation. The sequences of the component disaccharides were obtained from previous research^{30,31}. The

designations assigned to each disaccharide utilised in the investigation can be seen in Table 5.1.

Serotype	Disaccharide Structure	Designation	
		CHARMM	GLYCAM
19F	α -D-Glcp-(1 \rightarrow 2)- α -L-Rhap	G12R_C	G12R_G
19A	α -D-Glcp-(1 \rightarrow 3)- α -L-Rhap	G13R_C	G13R_G
Both	β -D-ManpNAc-(1 \rightarrow 4)- α -D-Glcp	N/A	M14G_G

Table 5.1. *Disaccharide designations*

The structural difference between the 19A and 19F serotype trisaccharide repeating units is seen in Table 5.1. In the 19A serotype the glycosidic linkage appears in the form of a 1-3 bond while in 19F a 1-2 bond is observed. The 1-2 linkage was designated G12R, while the 1-3 linkage was designated G13R. The M14G linkage was used in the construction of oligosaccharide extensions. A lack of availability of NAc group parameters for the CHARMM force field meant that the M14G disaccharide was only simulated in the GLYCAM force field.

The glucose primary alcohol group's O_{5'}-C_{5'}-C_{6'}-O_{6'} atoms were defined as the ω dihedral angle(Figure 5.1). This angle was used in all subsequent primary alcohol measurements.

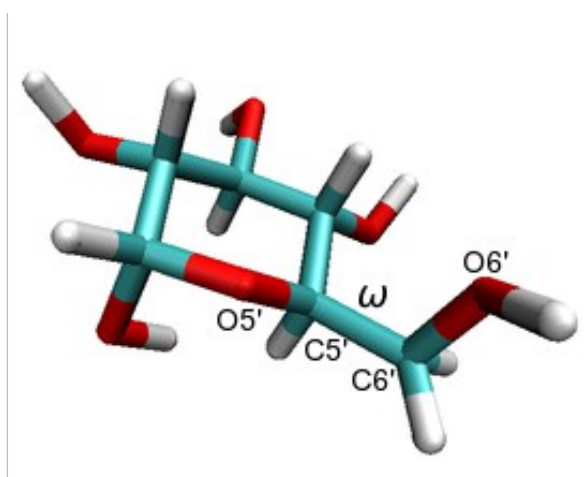


Figure 5.1. *Atoms defining the primary alcohol ω dihedral*

5.3. MD Simulation Parameters

In order to validate the biased simulations and extract representative conformations a series of unbiased simulations were performed on both the disaccharides of interest and glucose. These simulations were performed in a vacuum following a 10 000 step minimisation phase with a temperature control and equilibration regime involving 25K temperature reassignments culminating in a maximum temperature of 300K with simulation durations of 100ns. In order to better smooth the cutoff of electrostatic and van der Waal's forces, NAMD's switching algorithms were utilised with a cutoff distance of 15Å.

A point of concern does warrant specific mention. When using the default CHARMM force field in NAMD, the *scnb* value is not considered (*i.e.* 1-4 van der Waal's interactions are not scaled in the CHARMM force field). In a similar fashion to the CHARMM force field, GLYCAM requires no scaling of the 1-4 van der Waal's interactions. However the GLYCAM force field uses the AMBER file format, which requires the use of NAMD's AMBER parameters options. The use of the AMBER parameters option in turn involves the consideration of *scnb* values, which are defaulted to a value of 2. This is done to facilitate the requirements of traditional AMBER force fields. As this investigation called for equivalence in force field behaviours, 1-4 van der Waal's interaction scaling was prevented in the GLYCAM force field by using an explicit *scnb* value of 1.

For complete NAMD configuration file examples see appendix A.

5.4. Metadynamics Simulations

The Metadynamics sampling method was employed to generate accurate FESs for the disaccharides. The Metadynamics simulations were performed using the M14G, G12R_C, G12R_G, G13R_C and G13R_G disaccharides. Simulations were run for a duration of 1500ns. While 1500ns is usually an excessive duration for a MD simulation, the extended duration helped to ensure that the FESs of all disaccharides were fully explored.

Glycosidic ϕ and ψ dihedral angles were assigned as collective variables. The atoms used in defining the ϕ and ψ dihedral angles are shown in Table 5.2, with the positioning of the dihedrals shown in Figure 5.2. A bin width of 2.5° was employed with boundaries of -180° and 180° resulting in a representative FES allowing for full exploration of all possible glycosidic dihedral angle values (Appendix B).

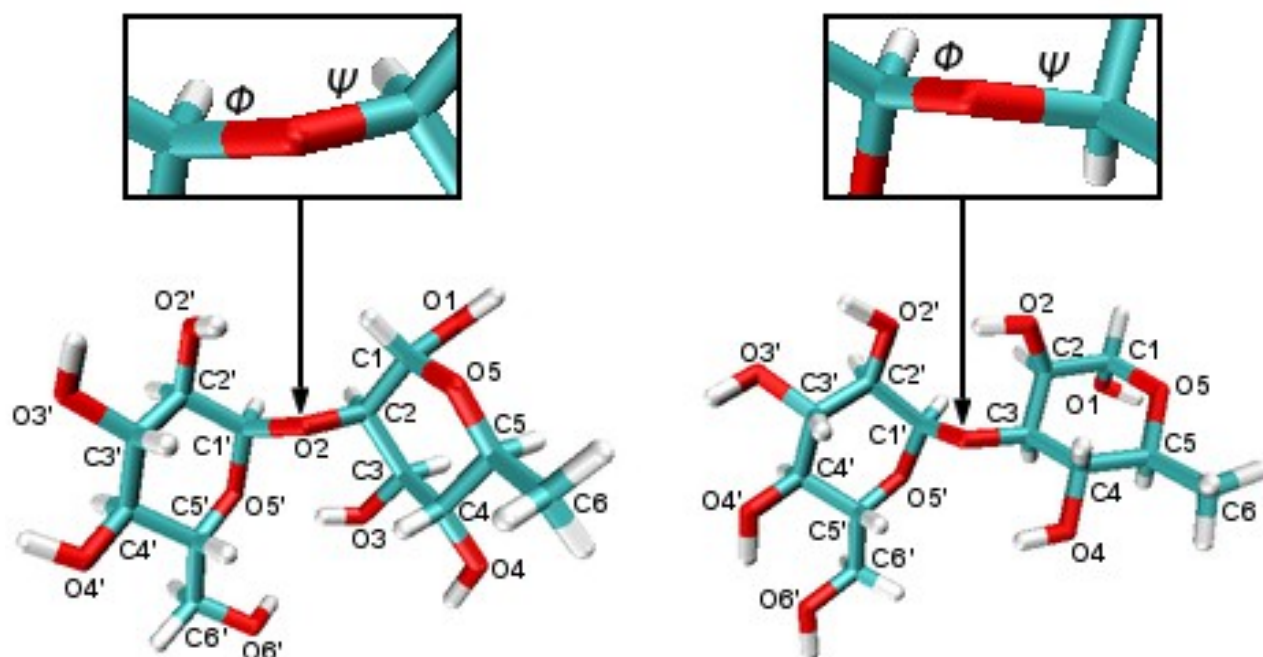


Figure 5.2. ϕ and ψ dihedral angles of G12R (left) and G13R (right)

Disaccharide	ϕ	ψ
G12R	H'1-C'1-O2-C2	C'1-O2-C2-H2
G13R	H'1-C'1-O3-C3	C'1-O3-C3-H3

Table 5.2. ϕ and ψ glycosidic dihedral angle component atoms

5.5. MD Analysis

Metadynamics sampling involves the use of the colvars module (Appendix B). The NAMD colvars module outputs the resultant FES in the form of a Potential of Mean Force (PMF)

file. This file consists of entries with Φ and Ψ angles along with the associated free energy value for the system. Using this PMF file, a 2D Ramachandran style FES²¹ with Φ , Ψ values on the X, Y axes respectively and free energy values depicted using coloured isocontour lines, was created for G12R_C, G12R_G, G13R_C and G13R_G.

The minima were determined and the corresponding frames of all disaccharides under all force fields were isolated and converted into trajectories. Analysis and comparison was then performed using the Visual Molecular Dynamics (VMD) application's Hbonds and Licorice graphical representation algorithms. A hydrogen bond limit of 3.4 Å at an angle of 100° was decided upon.

In order to better understand the likely transition paths between the energy wells, the lowest energy values for each Metadynamics Ψ bin were determined. This minimum energy path was then visualised as a plot of Ψ against free energy values and the differences between the minima and maxima along the path were determined.

Rotameric frequency figures were obtained to determine probability of occurrence of primary alcohol ω conformations. These probability figures were used to produce histograms identifying distinct peaks of high probability. The frames corresponding to those peaks were isolated allowing for analysis and comparisons across force fields.

5.6. Oligosaccharide Extensions

Using the in-house CarbBuilder software, three alternate oligosaccharide structures were built for both 19F and 19A sequences, to give a total of six molecules. Each oligosaccharide consisted of six trisaccharide repeating units. Minimum energy conformations obtained from the G12R, G13R and M14G FESs were used to set the dihedral angles in the oligosaccharides (Table 5.3).

Extension Designation	Force Field	Minimum Energy Conformation	Linkage
M14G13R_CA6	CHARMM	A	G13R
M14G13R_GA6	GLYCAM	A	G13R
M14G12R_CA6	CHARMM	A	G12R
M14G12R_CB6	CHARMM	B	G12R
M14G12R_GA6	GLYCAM	A	G12R
M14G12R_GB6	GLYCAM	B	G12R
ALL	GLYCAM	A	M14G
ALL			Phosphodiester

Table 5.3. *Oligosaccharide extension designations and component linkages*

The lack of NAc group parameters in the CHARMM force field meant that the global minimum of M14G under GLYCAM was utilised for all oligosaccharide extensions. Similarly, while CHARMM does possess some added phosphate parameters⁶⁶, both force fields lack phosphate parameters for the highly flexible anomeric phosphodiester linkages. This led to the usage of 0° for the phosphodiester torsion angles.

5.7. Software and Hardware

MD simulations were carried out using NAMD^{113,114} Linux-x86_64 version 2.9. NAMD was chosen as the MD application because of its ability to utilise multiple force field file formats, including in particular the CHARMM and AMBER formats. This feature allowed for a force field comparison within a single MD application thereby eliminating the potential impact of differences between MD applications.

MD simulations were performed on the University of Cape Town High Performance Computing Centre cluster nodes consisting of Intel Xeon 5660s running at 2800MHz. The cluster operated on Scientific Linux SL release 5.4 (Boron) with the use of the TORQUE Resource Management application to manage cluster resources.

The MacKerell group's CHARMM36 all-atom carbohydrate force field (April 2012 version) was used¹⁷. The CHARMM structure files (Protein Data Bank (PDB) and Protein Structure File (PSF)) were constructed using the In house CARBLOADER_builder.

For this investigation, GLYCAM 06h force field parameters were used¹⁰. The topology, coordinate and PDB files were constructed using the AMBER/GLYCAM Configurator tool also known as the Biomolecule Builder, available on the Woods Group's Glycam website⁷¹. The disaccharides were built by selecting the appropriate conformational isomer and residue information followed finally with an -OH sequence terminator.

The Visual MD (VMD) application version 1.9.1 and the accompanying Tcl scripting framework were utilised for the purposes of molecular visualisation and analysis. This included molecular visualisation, analysis of non-bonded interactions and creation of smaller, more manageable trajectories containing only the frames of interest (Appendix D).

Creation of graphical data representations was facilitated via the gnuplot command line plotting program, version 4.6 patch level 2.

5.8. Streptococcus pneumoniae Analysis Package

In addition to established applications such as NAMD and VMD, an in house Java application named the Streptococcus pneumoniae Analysis Package (StrAP) was written by the author during this investigation. The complete documented StrAP source code has been made available on GitHub¹¹⁵.

The long simulation durations required by the Metadynamics sampling methods resulted in unwieldy trajectories of 1.5 million frames. Written in Java, StrAP began as a handful of classes created to automatically search these large trajectories for frames expressing desired properties. However as the scope of the study increased, so too did the capabilities of StrAP. To date StrAP possesses the following features:

- **Fusing two separate VMD outputs into a single input file**
 - Using VMD frame lists and variable values are produced for each of the two marked variables in a trajectory. The lists must then be concatenated into a single file with the two datasets separated by a blank line prior to StrAP execution. StrAP then reads each line of the file transferring the contents into a String array. Once the second dataset is encountered StrAP then cross checks the frame numbers with the first dataset already present in the array and incorporates the second dataset's variable value into the array row of the appropriate frame.
- **Identifying trajectory frames expressing a specific dihedral angle or a pair of specific dihedrals (this feature is not limited to dihedrals and could be used to search for any kind of value that varies between frames, for example bond lengths)**
 - Taking in the fused (or a singular unfused list in the case of primary alcohol analyses) frame list StrAP rounds the angles in the list to whole numbers and then compares the first angle with the user defined desired angle. When a match is found a check on the second user defined angle is performed. If a match on the second angle is encountered then the corresponding frame number is outputted and the system continues to iterate through the input file.
- **Location of minima in FESs**
 - Taking in a PMF as input StrAP iterates through each line of the PMF looking for the lowest energy value. When a value lower than its currently recorded lowest value is encountered that value is stored as the lowest energy value. This continues until StrAP reaches the end of the file.
- **Identification of minimum energy paths through FESs**
 - StrAP iterates through the PMF file in 2.5, ψ increments/bins. For each increment it iterates through all corresponding ϕ values looking for the lowest energy value. Each energy value is compared against the recorded lowest energy value for that ψ bin. If a lower energy value is encountered it replaces the previous lowest energy value. This continues until all ϕ values have been iterated through at which point StrAP records the details of that lowest energy point in an output file and moves on to the next ψ increment. Once the final ψ

increment is processed the resulting output file contains the coordinates of the lowest energy points of every ψ increment.

- **Calculation of changes in free energy values between minimum and maximum values given a specified range of dihedral angles**
 - Given user defined ϕ and ψ ranges and a PMF file StrAP iterates through all ϕ and ψ coordinates within the range isolating the lowest and highest energy values. Once StrAP has iterated through the all coordinates within the range it outputs the highest and lowest energy values and the difference in energy between the two.
- **Calculation of frequency and probability of occurrence figures for any value that varies from frame to frame within a trajectory.**
 - StrAP creates an array in which the bins for ϕ and ψ angles and the amount of times each of those angle combinations are encountered are stored. Given a StrAP fused frame list as input, StrAP iterates through the list matching the ϕ and ψ angles in the list to entries in the array and incrementing the occurrence of that ϕ/ψ coordinate combination by 1. Once the final value on the input list has been processed the contents of the array are outputted to a file allowing for further analysis using graphing programs.

As a further aid for simplifying analysis, a Tcl script named `extractFrames`¹¹⁶ was written by the author for use within VMD. It extracts shorter, more usable trajectories from the original 1.5 million frame trajectory using StrAP's list of identified frames. The `extractFrames` script makes use of the `catdcd` executable written by Justin Gullingsrud and distributed under the University of Illinois Open Source License¹¹⁷.

In combination `extractFrames` and StrAP proved invaluable in the analysis of large simulation outputs.

Chapter 6: Results and Discussion

In this investigation first the results obtained for both the G12R and G13R disaccharides will be used to compare the CHARMM and GLYCAM force fields. The G12R and G13R disaccharides will then be compared directly. Finally, 6-unit polysaccharides assembled using the minimum energy conformations of the 19A and 19F component disaccharides will be compared in order to investigate the effect the G12R to G13R linkage shift has on serogroup 19 CPS conformations.

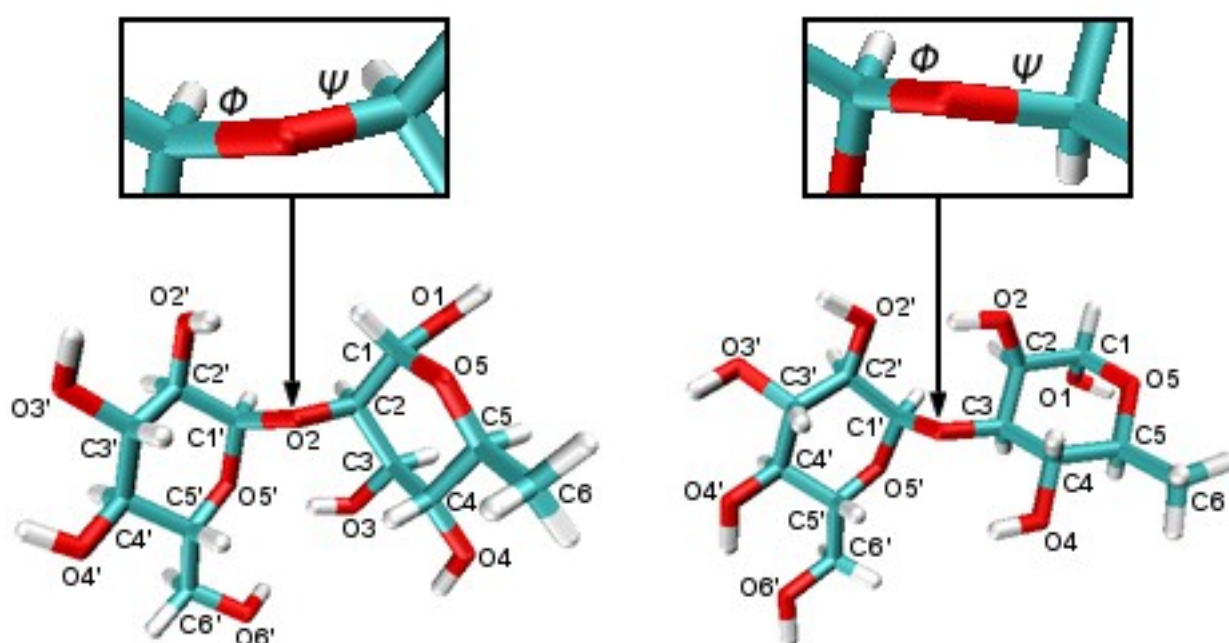


Figure 6.1. G12R (α -D-Glucose-(1-2)- α -L-Rhamnose) (left) and G13R (α -D-Glucose-(1-3)- α -L-Rhamnose) (right) disaccharides

The variation between the *S. pneumoniae* 19F and 19A serotypes is demonstrated by two key disaccharides: G12R and G13R (Figure 6.1). An investigation into the effects the shifting of the glycosidic bond from a 1-2 to a 1-3 conformation on the favourable conformations of the 19A and 19F polysaccharides could aid in future vaccine development.

The chief determinants of disaccharide conformations are the Φ and Ψ torsion angles for the glycosidic bond between residues (Figure 6.1). As these disaccharides are the key point of difference between 19A and 19F, a detailed analysis is a logical starting point.

6.1. Force Field Comparisons

As MD force fields have in the past yielded some disparate results when investigating disaccharides¹³, we compare the CHARMM and GLYCAM force fields. G12R and G13R were simulated using both force fields and the results are compared to determine whether or not any major differences could be observed in aspects such as overall FES, energy minima and relative free energy values, primary alcohol behaviour and hydrogen bonding patterns.

6.1.1. G12R (19F)

We begin with analysis of the α -D-Glucose-(1-2)- α -L-Rhamnose (G12R) disaccharide. CHARMM and GLYCAM Metadynamics simulations produced the FESs shown in Figure 6.2a and 6.2b, respectively.

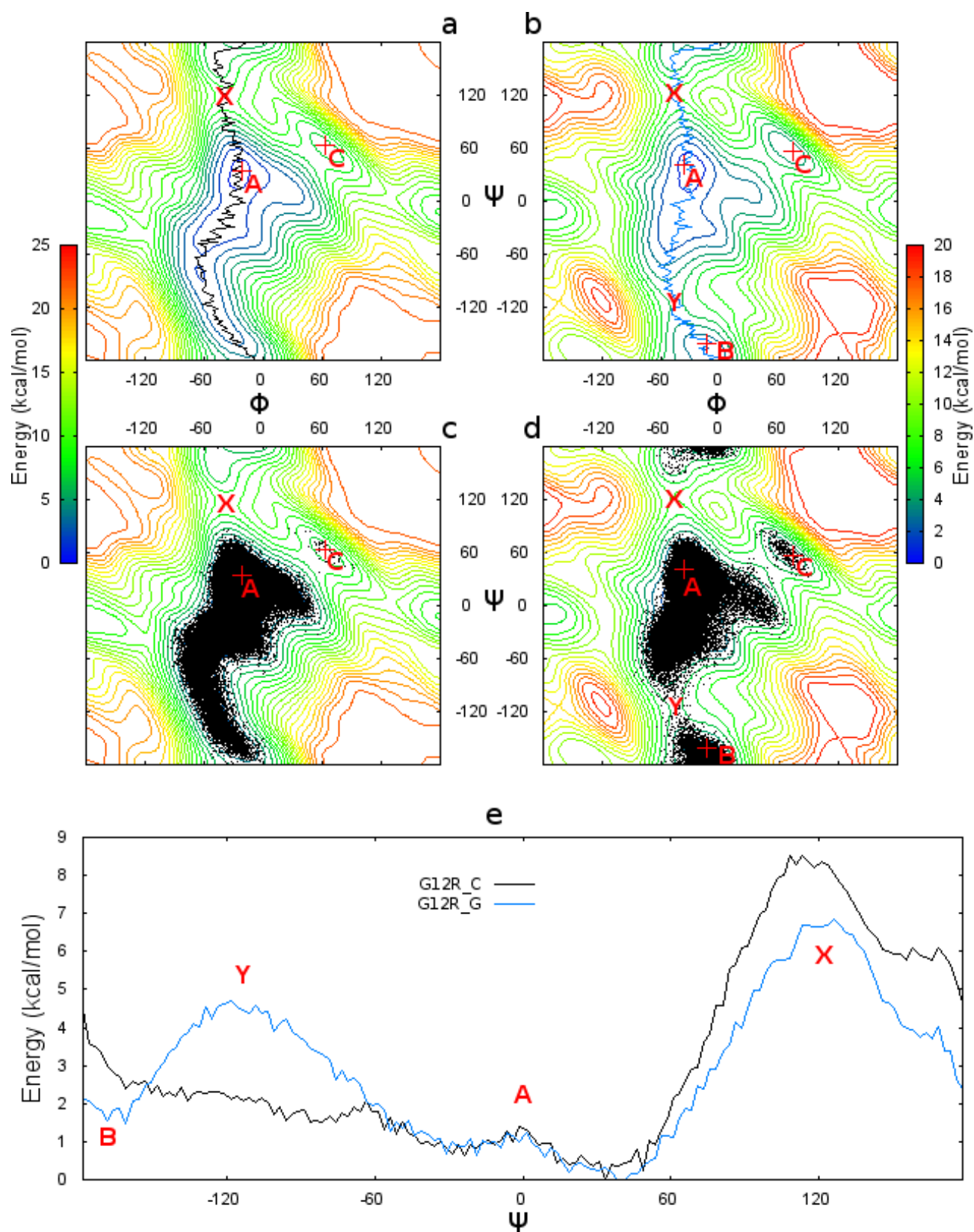


Figure 6.2. G12R_C (a), G12R_G (b) FESs. G12R_C (c) and G12R_G (d) unbiased scatter plots. Isocontour lines at 1kcal/mol intervals. Minimum energy paths (e), also shown in (a) and (b).

	Minima/Barriers (Φ ; Ψ ; $\Delta G(\text{kcal/mol})$)				
	A	B	C	X	Y
CHARMM	-21.25°; 33.75°; 0.0	-41.25; -136.25; 2.08	63.75°; 63.75°; 5.87	-33.75°; 113.75°; 8.54	-
GLYCAM	-36.25°; 41.25°; 0.0	-13.75°; -161.25°; 1.48	73.75°; 56.25°; 2.61	-51.25°; 126.25°; 6.86	-38.75°; -118.75°; 4.72
MM2 ⁴	-34°; -35°; 0.0	-	-	-	-
MM ⁺ 45	-35°; -38°; 0.0	-	-	-	-

Table 6.1. *G12R minima and barrier torsion angles*

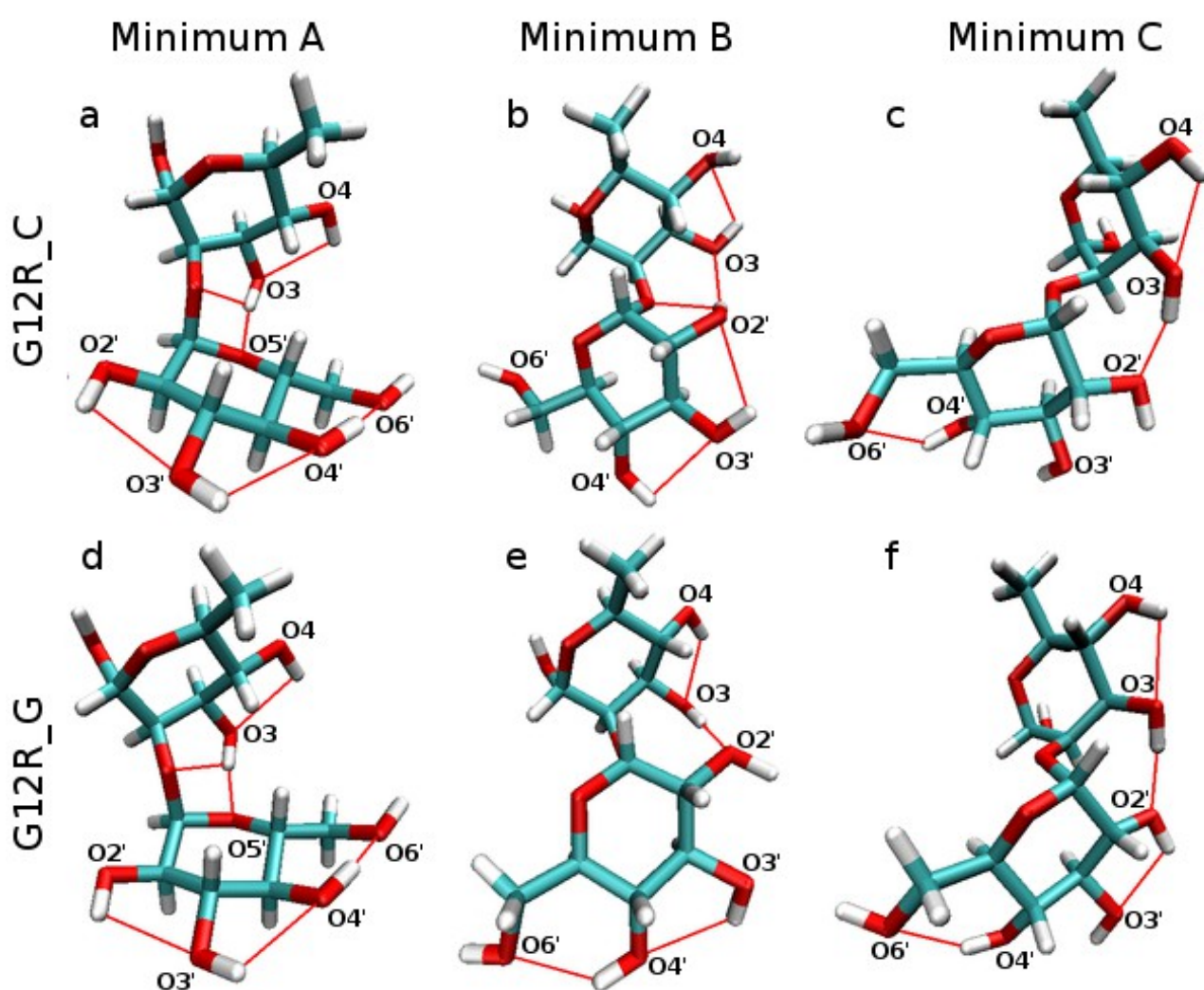


Figure 6.3. *G12R_C (a, b and c) and G12R_G (d, e and f) minima. The red lines are hydrogen bonds.*

With extensive favourable areas at or below 4kcal/mol the G12R FESs (Figure 6.2) for the CHARMM (left column) and GLYCAM (right column) force fields, both predict a flexible dimer. The force fields also agree on the general location of the G12R global minimum energy conformation in the central well (A) (Φ , Ψ = -21.25°, 33.75° under CHARMM and -36.25°; 41.25° under GLYCAM (Table 6.1)). Under both force fields this minimum is characterised by minimal steric strain, a wealth of intraresidue hydrogen bonds and a stabilising O3-O'5 interresidue hydrogen bond (Figure 6.3). While hydrogen bonds are believed to play a role in structural stabilisation of carbohydrates⁵, their effectiveness is likely to decrease in an aqueous environment¹¹⁸. The global minima are in disagreement with previous global minima as determined using the MM2 (Φ , Ψ = -34°, -35°)⁴ and MM⁺ (Φ , Ψ = -35°, -38°)⁴⁵ force fields both of which favoured the negative Ψ region of the A well.

The anti- Ψ region of the G12R FES (B in Figure 6.2) exhibits the major difference between the CHARMM and GLYCAM force fields' G12R predictions. Under GLYCAM a distinct energy well is present with a secondary minimum B (Φ , Ψ = -13.75°, 41.25° (Table 6.1)). Under CHARMM wells A and B are merged, with well A forming a long valley extending into the B region (though not as far as GLYCAM's B well which extends into positive- Ψ). This results in the lack of a Y barrier and no distinct B well. Instead a conformation of local low energy is observed towards the end of the valley (Φ , Ψ = -41.25°, -136.25° (Table 6.1)). Under GLYCAM aliphatic hydrogens carry a net charge of 0, while under CHARMM a charge of +0.09 is used^{10,119}. This may explain the discrepancies. Conformations in the Y region bring H1 and O'5, and H'1 and O5 within 3Å of one another (Figure 6.4). CHARMM's aliphatic charges may make these pairwise interactions stronger, lowering the energy of Y region conformations. Conversely the close quarters between the rhamnose and glucose residues in the B region brings aliphatic hydrogens into close proximity (Figure 6.3). Resulting H'1-H4 and H'5-H1 pairwise interactions are likely stronger under CHARMM, increasing the energy of the B region conformations. A more comprehensive investigation is required.

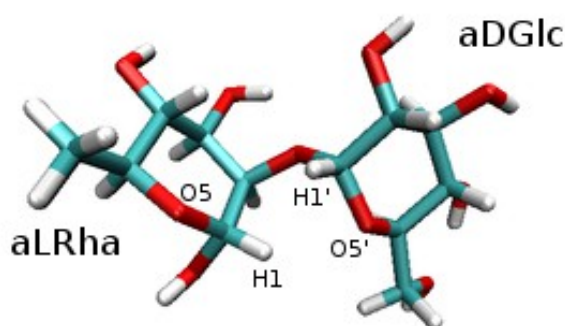


Figure 6.4. G12R Y region conformation (Φ , Ψ = -44° , -116°)

A tertiary minimum C (Φ , Ψ = 63.75° , 63.75° and 73.75° , 56.25° under CHARMM and GLYCAM respectively (Table 6.1)) was identified by both force fields. Well C is of far lower energy under GLYCAM (2.61kcal/mol) than under CHARMM (5.87kcal/mol) and is characterised by an O3-O'2 interresidue hydrogen bond under both force fields (Figure 6.3).

The unbiased 100ns simulation scatter plot superimposed onto the FESs show that barrier X (ΔG = 8.54kcal/mol under CHARMM and ΔG = 6.86kcal/mol under GLYCAM (Table 6.1)) is too high in energy to be explored in either force field (Figure 6.2c and d). As expected with the lower energy of the region under GLYCAM, the C well shows more frequent occupation under GLYCAM than CHARMM. The effect of the differences between the force fields' A and B regions is highlighted. The CHARMM FES shows heavy occupation of the extended A well while GLYCAM shows occupation of the A and B wells with slight occupation of the Y region indicating that it is the path of transition. The path of minimum energy (Figure 6.2e) further highlights the long CHARMM valley with absent Y peak.

We now move on to an analysis of G12R primary alcohol behaviour. CHARMM conformations corresponding to GLYCAM's global minimum (Φ , Ψ = -36.25° , 41.25°), involve close proximity between O3 and O'5. The O3-O'5 interresidue hydrogen bond persists (Figure 6.5). However, when in the g^+ conformation, the glucose residue's primary alcohol group approaches the O3-O'5 bond, this appears to result in a shift in the relative orientations of O3 and O'2 creating a pair of potential hydrogen bonds: O3-O'5 and O3-O'6 (Figure 6.5d). At $\sim 113^\circ$ and $\sim 144^\circ$ respectively, both of these bonds are further from ideal than the original O3-O'5 hydrogen bond seen in the G12R_C minimum (angle, distance =

173.22°, 2.8Å). This shift is not observed to occur on the same scale at Φ ; $\Psi = -21.25^\circ$; 33.75°. It is possible that CHARMM models the single hydrogen bond prevalent in its minimum conformation as more favourable than the O3-O'5 and O3-O'6 pair resulting a shifting of the global minimum coordinates.

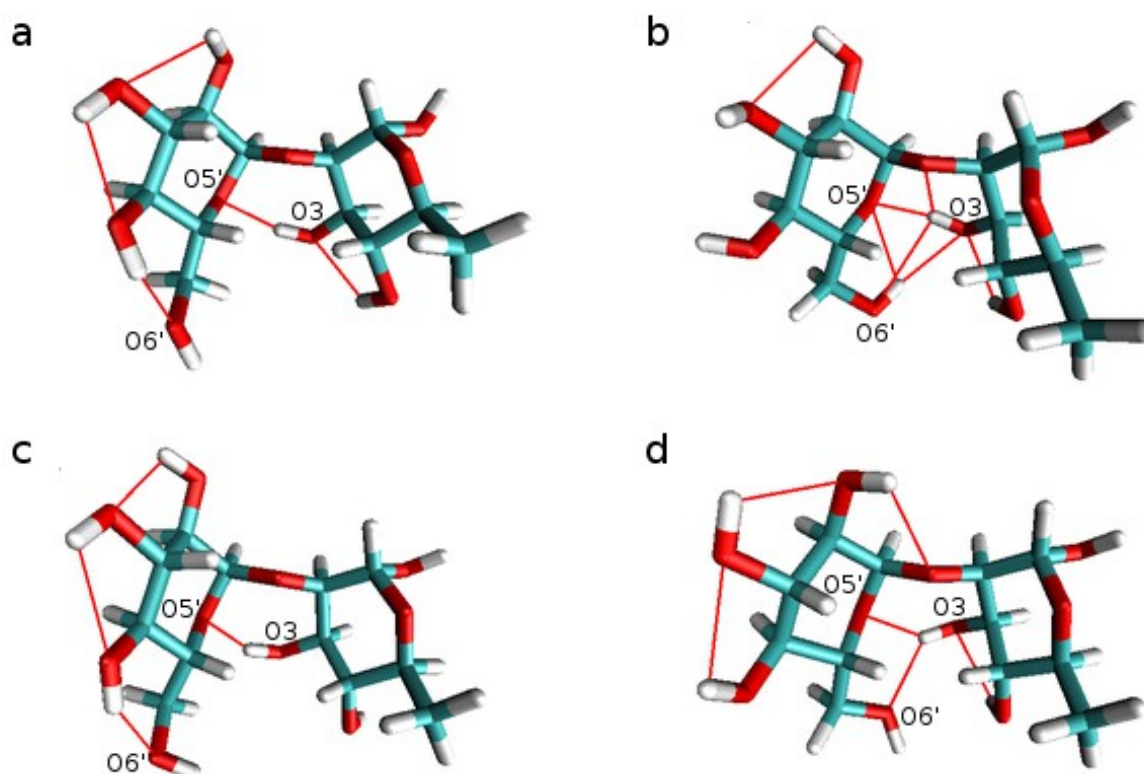


Figure 6.5. *G12R_C Conformations showing the interference by O'6 with the O₃-O'5 hydrogen bond. (a): no disruption (Φ , $\Psi = -21.25^\circ$; 33.75°). (b): ineffective disruption (Φ , $\Psi = -21.25^\circ$; 33.75°). (c): no disruption (Φ , $\Psi = -36.25^\circ$, 41.25°). (d): effective disruption (Φ , $\Psi = -36.25^\circ$, 41.25°). The red lines indicate hydrogen bonds.*

Under GLYCAM the G12R disaccharide appears to not experience the paired interaction at Φ , $\Psi = -36.25^\circ$, 41.25° on the scale seen in the CHARMM force field. Where the interaction would occur under CHARMM, the hydroxyls are instead orientated to allow a single O3-O'6 bond without the intermediate orientations seen under CHARMM.

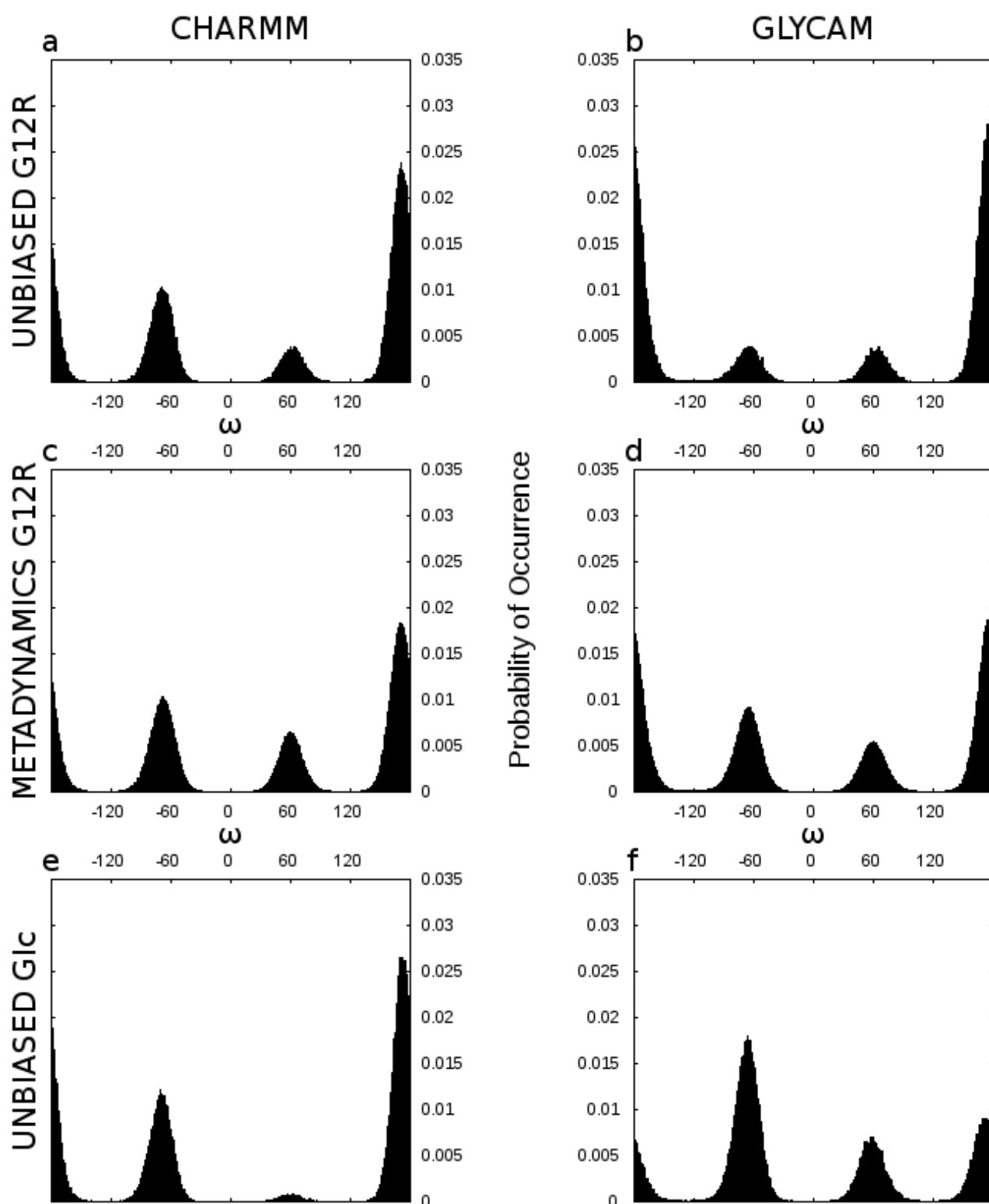


Figure 6.6. Probability of occurrence of G12R and Glc primary alcohol $O_5-C_4-C_6-O_6$ ω dihedral angles

It is expected that the glucose primary alcohol group will tend to favour 3 traditional conformations with respect to the $O_5-C_5-C_6-O_6$ ω dihedral angle in order to reduce steric

strain. These conformations are trans ($\omega = 180$), gauche⁺ (g⁺) ($\omega = 60$) and gauche⁻ (g⁻) ($\omega = -60$). As can be seen from Figure 6.6, G12R is no exception, with 3 distinct peaks around $\omega = 180$, $\omega = 60$ and $\omega = -60$ for both force fields.

	Conformation (ω ; probability)		
	Unbiased Simulations		
	trans	gauche ⁺	gauche ⁻
Glc_C	172°; 0.02654	62°; 0.0008	-71°; 0.01218
Glc_G	173°; 0.00901	58°; 0.00702	-66°; 0.01794
G12R_C	171°; 0.02383	63°; 0.00389	-69°; 0.01032
G12R_G	177°; 0.02845	58°; 0.00362	-65°; 0.00377
	Biased Simulations		
	trans	gauche ⁺	gauche ⁻
G12R_C	170°; 0.01833	59°; 0.00647	-68°; 0.01025
G12R_G	175°; 0.01867	60°; 0.00542	-65°; 0.00914

Table 6.2. Peak conformation probabilities of the O₅-C₅-C₆-O₆ ω dihedral angle for the unbiased Glc, biased G12R and unbiased G12R simulations

Table 6.2 shows the highest values of each peak seen in Figure 6.6. With the exception of glucose in isolation, in which GLYCAM shows similar affinity for all conformations, both force fields favour the trans conformation for G12R. The g⁻ conformation occurs less frequently than trans, often with less than half the probability of occurrence with the least popular being g⁺. In CHARMM glucose in isolation heavily favours the trans conformation. The heavy favouring of the trans conformation does not fit with experimental evidence and has been addressed in previous work by Kuttel *et al.* through the CSFF²⁰. Considering the attention given to CHARMM parametrisation for solution, the favouring of trans may be lessened in an aqueous environment.

The orientation of the primary alcohol group affects both the O'4-O'6 intraresidue and O3-O'6 interresidue hydrogen bonds: in g⁻ neither bond is possible; in g⁺, the O'4-O'6 bond is not possible but the O3-O'6 bond is; in trans conformation the O'4-O'6 bond is possible but the O3-O'6 is not (Figure 6.7). The lack of a strong intraresidue O'4-O'6 hydrogen bond

may account for the lack of gauche favourability while the additional presence of the seemingly unfavourable O3-O'6 interaction in the g^+ conformation could explain its further decreased popularity. Conversely the presence of the O4-O'6 bond and the absence of O3-O'6 interaction may address the favouring of the trans conformation.

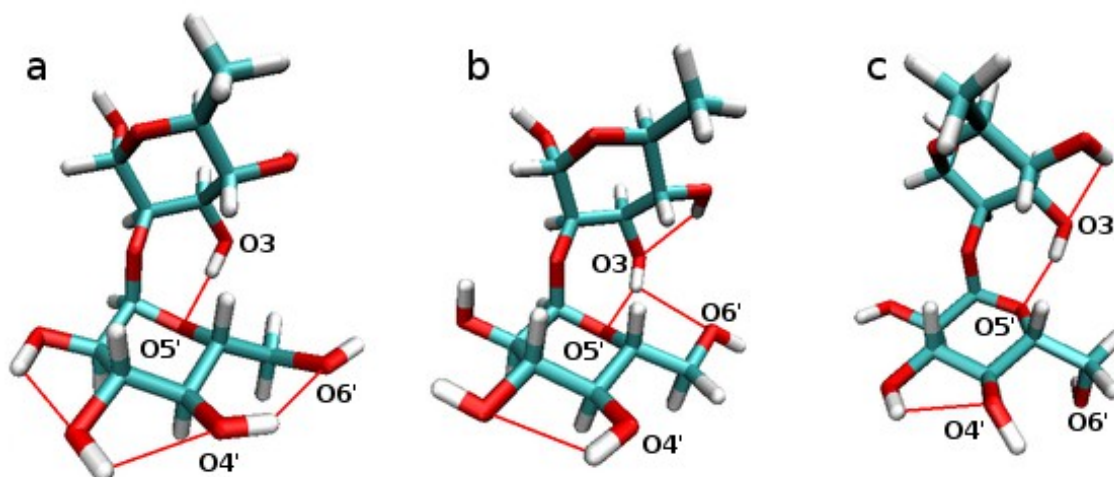


Figure 6.7. *G12R_C* primary alcohol in trans (a), g^+ (b) and g^- (c). The red lines indicate hydrogen bonds

6.1.2. G13R (19A)

The investigation moves now to a comparative analysis of CHARMM and GLYCAM results for the α -D-Glucose-(1-3)- α -L-Rhamnose (G13R) disaccharide. As with G12R, the investigation begins with FESs generated using the ϕ and ψ torsion angles of the G13R glycosidic linkage as collective variables in a Metadynamics sampling technique (Figure 6.8). ϕ and ψ torsion angle coordinates visited during an unbiased simulation were collated into a scatter plot and superimposed over the FESs (Figure 6.8c and 6.8d) in order to provide a general overview of the nature of the G13R disaccharide.

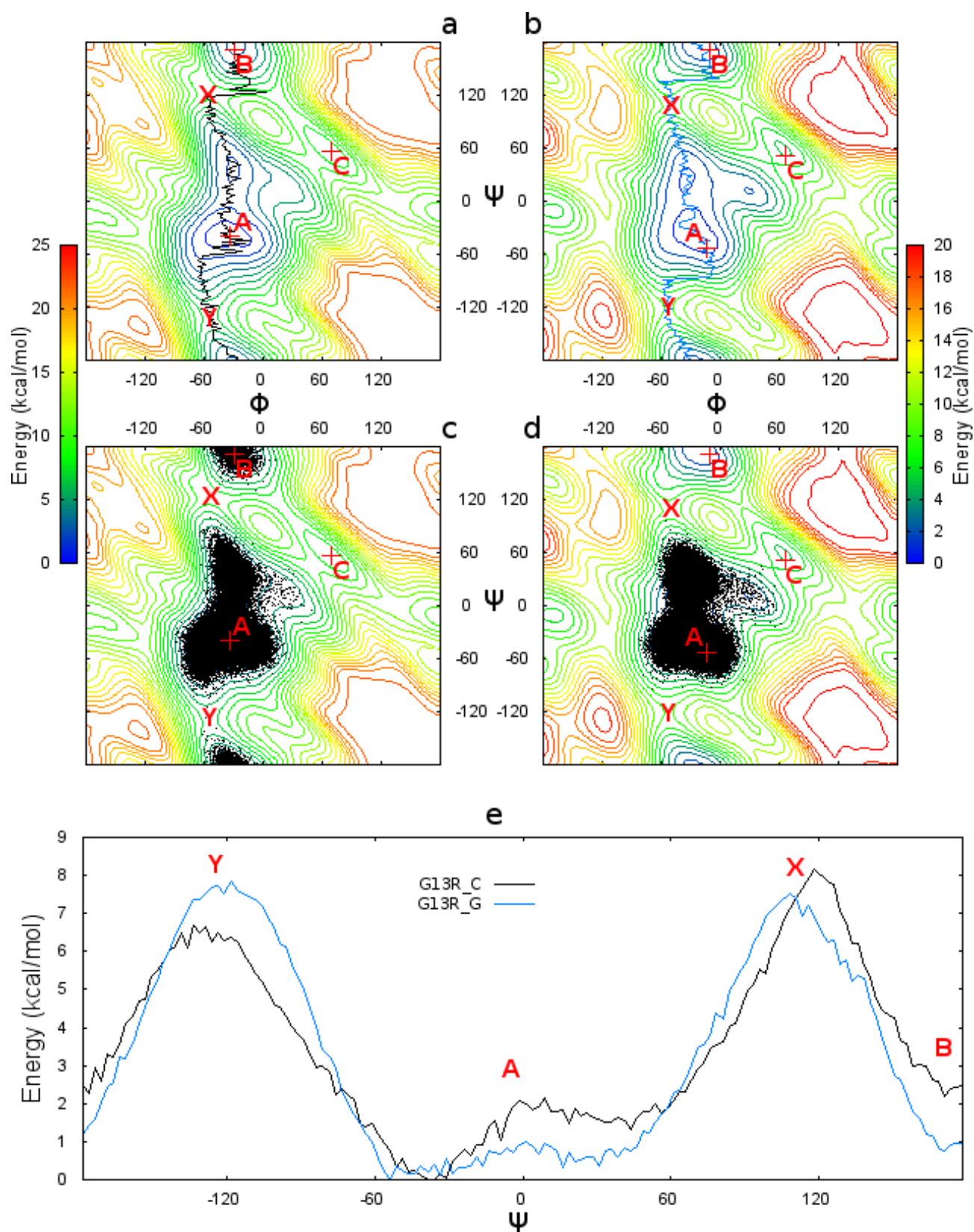


Figure 6.8. G13R_C (a), G13R_G (b) FESs. G13R_C (c) and G13R_G (d) unbiased scatter plots. Isocontour lines at 1kcal/mol intervals. Minimum energy paths (e), also shown in (a) and (b).

	Minima/Barriers (Φ ; Ψ ; $\Delta G(\text{kcal/mol})$)				
	A	B	C	X	Y
CHARMM	-33.75°; -38.75°; 0.0	-28.75°; 171.25°; 2.18	68.75°; 56.25°; 6.67	-58.75°; 118.75°; 8.16	-51.25°; -133.75°; 6.68
	-28.75°; 43.75°; 1.31				
GLYCAM	-13.75°; -53.75°; 0.0	-11.25°; 171.25°; 0.78	66.25°; 51.25°; 4.18	-53.75°; 108.75°; 7.53	-56.25°; -118.75°; 7.84
	36.25°; 16.25°; 0.32				
	28.75°; 6.25°; 1.35				
MM2 ⁴	-35°; -38°; 0.0	-28°; -175°; 4.3	-		

Table 6.3. *G13R minima and barrier torsion angles*

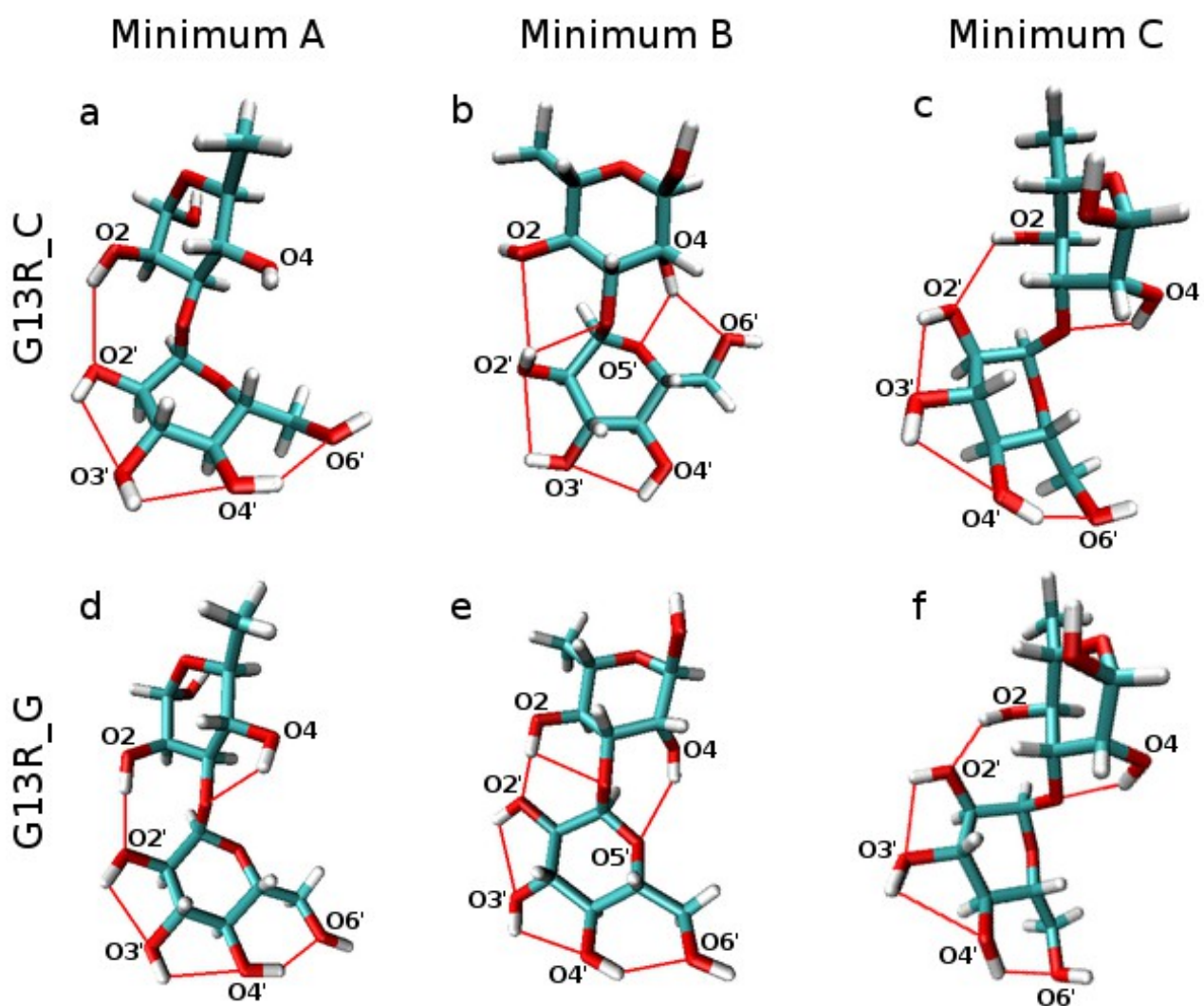


Figure 6.9. *G13R minima conformations. The red lines indicate hydrogen bonds*

The G13R FESs (Figure 6.8) for CHARMM (left) and GLYCAM (right), are more in agreement than was the case in G12R. In both force fields the central A well contains the global minimum ($\Phi, \Psi = -33.75^\circ, -38.75^\circ$ for CHARMM and $\Phi, \Psi = -13.75^\circ, -53.75^\circ$ for GLYCAM (Table 6.3)) characterised by an O2-O'2 interresidue hydrogen bond (Figure 6.9). Both force fields show two regions of low energy conformations, a favoured negative- Ψ region containing the global minimum and a secondary positive- Ψ region. The A well is broader with lower energy states than seen under CHARMM: GLYCAM exhibits a less favoured tertiary positive- Φ region ($\Phi, \Psi = 28.75^\circ, 6.25^\circ$ (Table 6.3)) absent in CHARMM and a secondary local minimum of lower energy than that of CHARMM ($\Phi, \Psi, \Delta G = -28.75^\circ, 43.75^\circ, 1.31\text{kcal/mol}$ for CHARMM and $\Phi, \Psi, \Delta G = -36.25^\circ, 16.25^\circ, 0.32\text{kcal/mol}$ for GLYCAM (Table 6.3)). Previous HSEA and MM2 force field calculations presented a global minimum of $\Phi, \Psi = -35^\circ, -38^\circ$ ⁴ which, while in close agreement with both force fields, is particularly close to the CHARMM global minimum.

Unlike in G12R, both force fields predict a distinct well in region B containing a local minimum. Located at $\Phi, \Psi = -28.75^\circ, 171.25^\circ$, for CHARMM and $\Phi, \Psi = -11.25^\circ, 171.25^\circ$, for GLYCAM (Table 6.3), the minimum is of considerably lower energy under GLYCAM (0.78kcal/mol under GLYCAM and 2.18kcal/mol under CHARMM (Table 6.3)). The previous MM2 investigation identified a tertiary minimum in the B region at $\Phi, \Psi = -28^\circ, -175^\circ$, though at 4.3kcal/mol this minimum is of significantly higher energy⁴.

The well in region C is similarly positioned in both CHARMM ($\Phi, \Psi = 68.75^\circ, 56.25^\circ$) and GLYCAM ($\Phi, \Psi = 66.25^\circ, 51.25^\circ$) (Figure 6.8 and Table 6.3). Characterised by an O2-O'2 hydrogen bond (Figure 6.9), GLYCAM (4.18kcal/mol) predicts a lower energy minimum than that of CHARMM (6.67kcal/mol) (Table 6.3).

Under GLYCAM the Y (7.84kcal/mol) and X (7.53kcal/mol) barriers have similar heights (Figure 6.8 and Table 6.3) while under CHARMM the Y barrier (6.68kcal/mol) is lower than the X barrier (8.16kcal/mol). The higher Y values make transition more difficult and, as seen in the scatter plots of the unbiased 100ns simulations (Figure 6.8c and d), transition from well A to B occurs only under CHARMM, though the sparse population of CHARMM's Y region suggests that transition is less frequent than that observed for G12R. GLYCAM

well A's tertiary energy well is more densely populated than the equivalent area in CHARMM while well C remains unexplored under both force fields (Figure 6.3).

The behaviour of the primary alcohol group in G13R is now further investigated. Figure 6.10 and Table 6.4 document the overall probability of occurrence of primary alcohol group O5-C5-C6-O6 ω torsion angles in the biased and unbiased G13R MD simulations of both force fields.

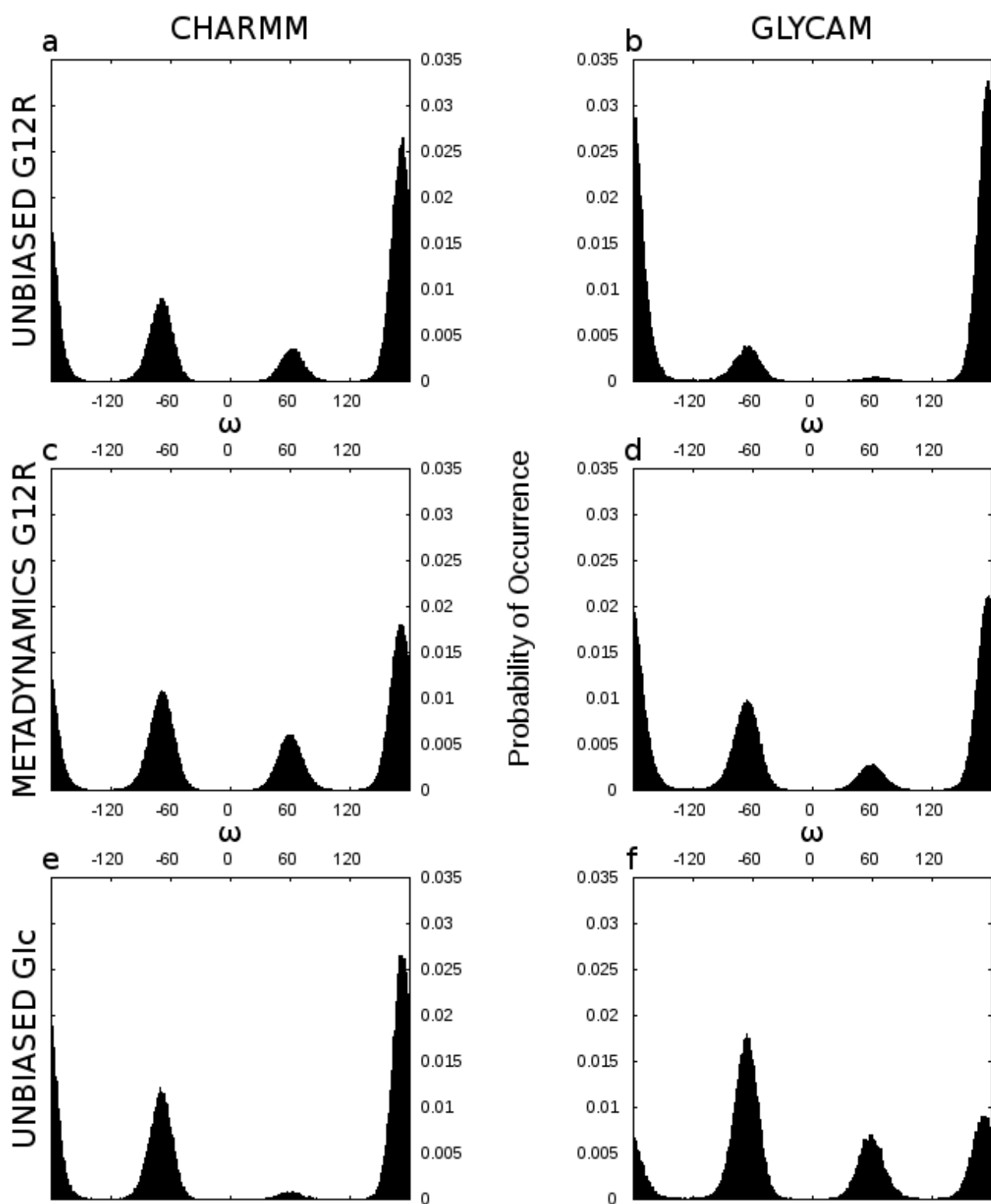


Figure 6.10. Probability of occurrence of G12R and G1c primary alcohol O5-C4-C6-O6 ω dihedral angles

	Conformation (ω; frequency)		
	Unbiased Simulations		
	trans	gauche⁺	gauche⁻
Glc_C	172°; 0.02654	62°; 0.0008	-71°; 0.01218
Glc_G	173°; 0.00901	58°; 0.00702	-66°; 0.01794
G13R_C	173°; 0.02657	60°; 0.00347	-68°; 0.00903
G13R_G	176°; 0.03271	63°; 0.00047	-68°; 0.00379
	Biased Simulations		
	trans	gauche⁺	gauche⁻
G13R_C	171°; 0.01805	61°; 0.00601	-69°; 0.01088
G13R_G	177°; 0.02112	60°; 0.00278	-66°; 0.00979

Table 6.4. Peak conformation probabilities of the O'₅-C'₅-C'₆-O'₆ ω dihedral angle for the unbiased Glc, biased G13R and unbiased G13R simulations

In G13R the trans, g⁺ and g⁻ peaks are at 171°, 61° and -69° under CHARMM and 177°, 60° and -66° in GLYCAM respectively (Table 6.4). As in G12R, with the exception of GLYCAM's 100ns glucose simulation, both force fields favour the trans conformation. The lower g⁻ probability seen in unbiased G13R_G simulations occurs only in that simulation and may be a result of the largely unexplored FES stemming from the lack of transition from well A to well B (Figure 6.8d). There is also a noticeable difference in the probability of occurrence of the g⁺ conformation between force fields. GLYCAM produces g⁺ frequencies significantly lower than those of CHARMM in all simulations (Table 6.4).

The G13R primary alcohol group's conformation influences both the O4-O'6 and O'4-O'6 intraresidue hydrogen bond: in trans the intraresidue bond is retained but the interresidue bond is not; in g⁺ the interresidue bond is possible but the intraresidue bond is not; in g⁻ neither bond can be formed (Figure 6.11). The favouring of trans may be due to its orientation allowing for the formation of the intraresidue O'4-O'6 hydrogen bond. As the impact of hydrogen bonds has been known to change in aqueous solutions¹¹⁸ it is possible that the favouring of trans is a phenomenon isolated to vacuum conditions.

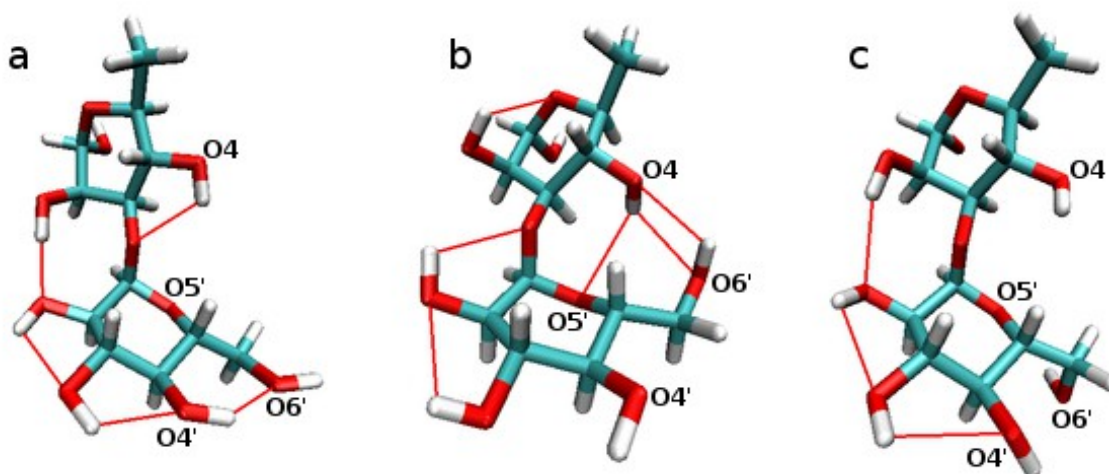


Figure 6.11. *G13R_C* primary alcohol in *trans* (a), *g⁺* (b) and *g⁻* (c) conformations. The red lines indicate hydrogen bonds

Despite some discrepancies, in both the G12R and G13R the FESs obtained with both force fields are in close agreement in terms of location of minimum energy conformations. GLYCAM consistently predicts broader energy wells with local minima of lower values than those of CHARMM. It can therefore be said that GLYCAM predicts an altogether more flexible disaccharide than CHARMM with less potential for well transition.

6.2. Linkage Comparison

G12R and G13R differences have implications for the conformation of the serogroup 19 trisaccharide repeating units and the CPSs those repeating units form. A small change in the most favoured glycosidic ϕ and ψ dihedral angles could result in considerable conformational differences when compounded over a chain of repeating units.

The MM2 study predicted similar G12R and G13R FESs. With its single centralised well of low energy, the FESs described disaccharides of limited flexibility⁴. The application of modern day force fields orientated specifically towards the simulation of carbohydrates has improved upon this view. The CHARMM and GLYCAM simulations have predicted more flexible disaccharides in both G12R and G13R. While both G12R and G13R have similar ϕ and ψ ranges (Figure 6.12), G12R appears to be the more flexible disaccharide with

frequently explored anti- ψ conformations (Figure 6.2c and d). Both force fields show a rise in X and Y barrier heights for G13R over those of G12R impeding rotation to anti- ψ conformations in G13R (Figure 6.13). This is more noticeable under CHARM where G12R lacks a distinct Y barrier and G13R possesses a sizeable 6.68kcal/mol Y barrier. Under GLYCAM the trend is less pronounced with the Y barrier rising from 4.72kcal/mol in G12R to 7.84kcal/mol in G13R. Further reinforcing this observation is the seemingly rare transition between wells under CHARMM and the complete absence of transition under GLYCAM in G13R (Figure 6.8c and d).

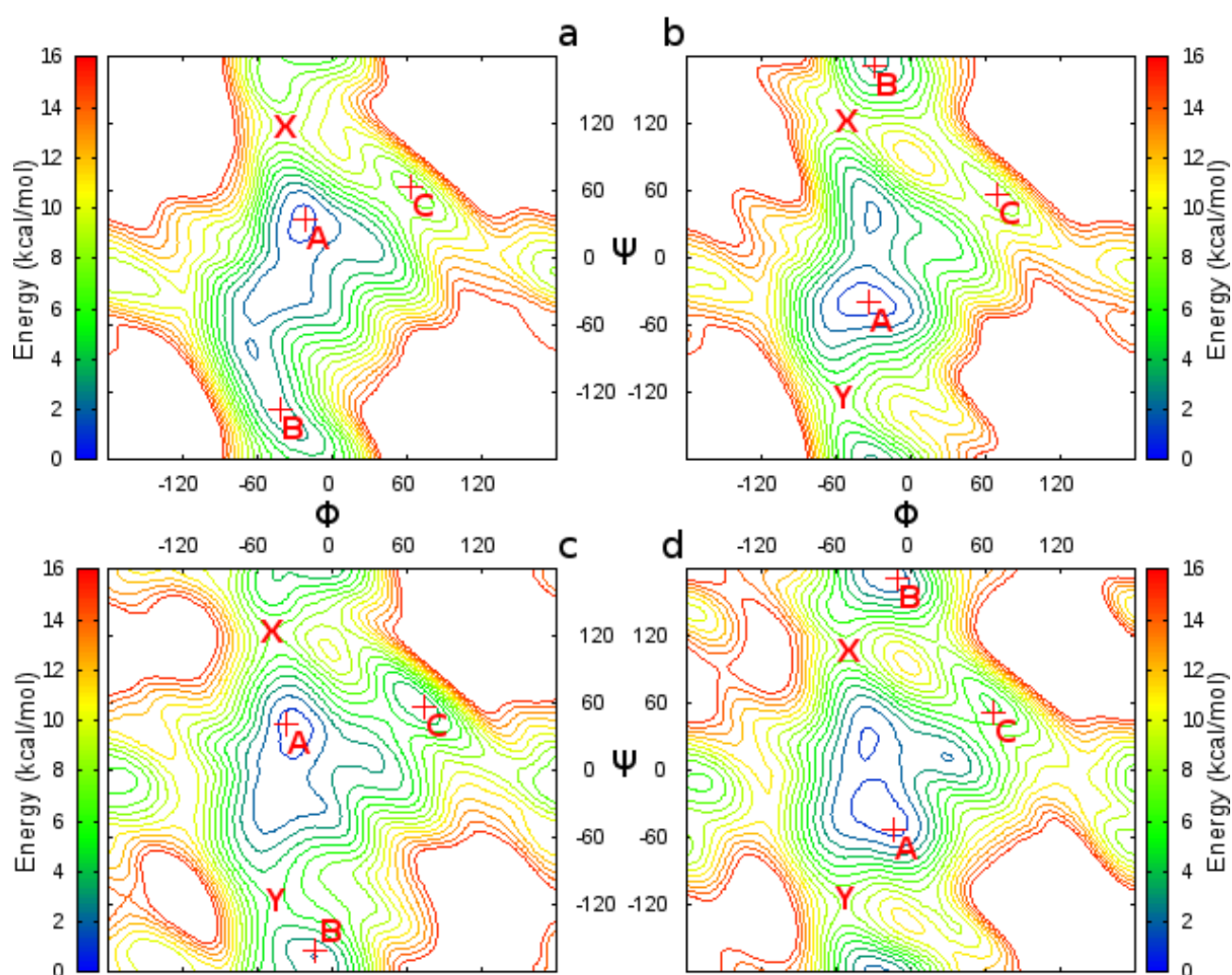


Figure 6.12. FESs of G12R_C (a), G12R_G (c), G13R_C (b) and G13R_G (d). The isocontour lines are 1kcal/mol increments with a 15kcal/mol cutoff

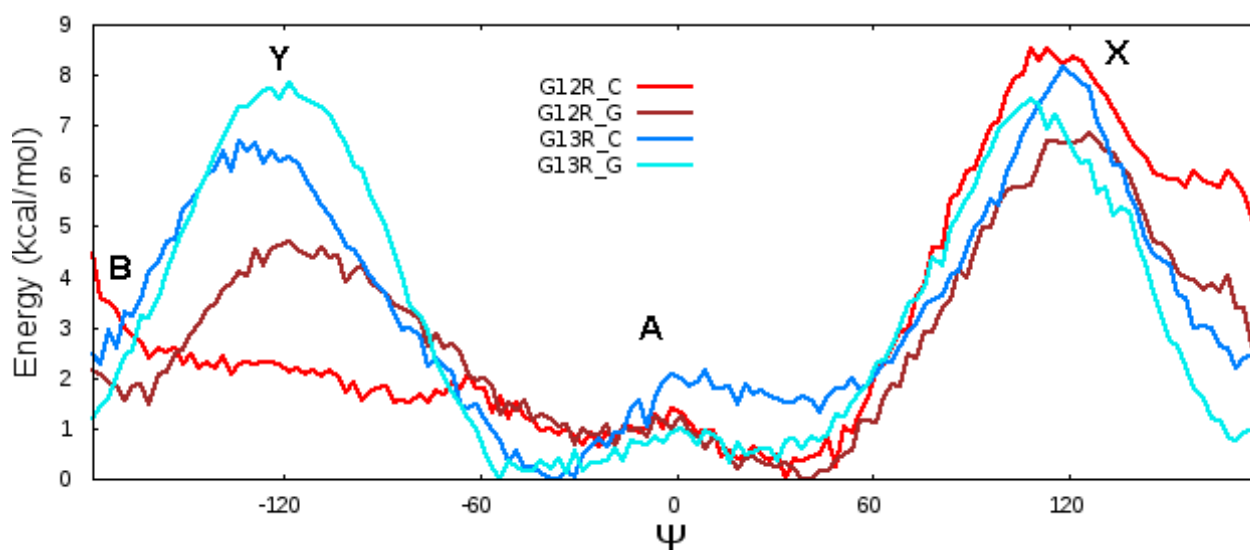


Figure 6.13. Minimum energy paths for G12R and G13R. Key wells and peaks from Figure 6.12 are labelled

The FESs contain some differences, the most noticeable being the merging of G12R wells A and B under CHARMM while G13R exhibits a distinct secondary energy well B under both force fields (Figure 6.12). In a similar manner the global minima torsion angles differ dramatically in response to the (1-2) to (1-3) glycosidic shift. While ϕ angles show little change the ψ angles shift by 72.5° under CHARMM and a remarkable 95° under GLYCAM (Table 6.6).

Force Field	G12R Minimum (ϕ ; ψ)	G13R Minimum (ϕ ; ψ)	Degree Shift (ϕ ; ψ)
CHARMM	-21.25° ; 33.75°	-33.75° ; -38.75°	12.5° ; 72.5°
GLYCAM	-36.25° ; 41.25°	-13.75° ; -53.75°	22.5° ; 95°

Table 6.5. Degree shift in G12R and G13R global minima

6.3. Oligosaccharide Extensions

The oligosaccharide repeating unit chains depicted in this section are limited in their application due to the lack of availability of reliable phosphodiester bond parameters for both CHARMM and GLYCAM. While potential conformational impacts are highlighted, inferences onto actual *in vivo* 19A and 19F CPSs are not necessarily reliable.

Extension	Force Field	Minimum Energy Conformation	Linkage	Dihedral (Φ ; Ψ)
M14G13R_CA6	CHARMM	A	G13R	-34°; -39°
M14G13R_GA6	GLYCAM	A	G13R	-14°; -54°
M14G12R_CA6	CHARMM	A	G12R	-21°; 34°
M14G12R_CB6	CHARMM	B	G12R	-41°; -136°
M14G12R_GA6	GLYCAM	A	G12R	-36°; 41°
M14G12R_GB6	GLYCAM	B	G12R	-14°; -161°
ALL	GLYCAM	A	M14G	49°; -14°
ALL			Phosphodiester	0°; 0°

Table 6.6. *Oligosaccharide extension designations*

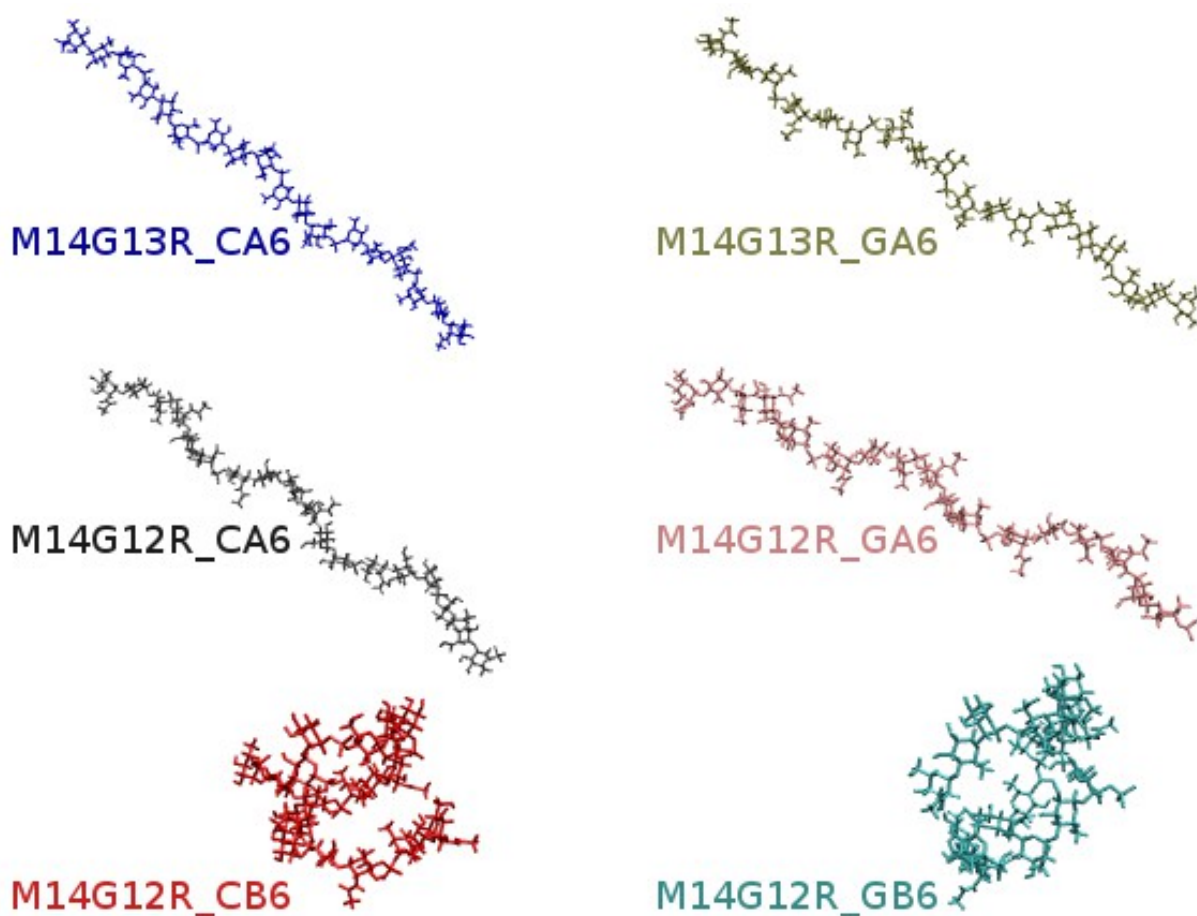


Figure 6.14. 6-unit oligosaccharide extension structures

Overall the extensions utilising G12R or G13R global minimum energy conformations (M14G13R_CA6, M14G13R_GA6, M14G12R_CA6 and M14G12R_GA6 (Table 6.6)) show few structural discrepancies (Figure 6.14). While the M14G12R_CA6 helix is slightly more tightly coiled, all four extensions exhibit a flat extended helical structure regardless of force field or component disaccharides utilised. In the extensions utilising G12R or G13R global minimum energy conformations, the residues and the phosphate group remain readily accessible for antibody binding. The minor conformational variations are therefore unlikely to explain the lack of 19F to 19A cross protection.

The extensions utilising G12R B minimum energy conformations under CHARMM (M14G12R_CB6) and GLYCAM (M14G12R_GB6) differ structurally from the global minima extensions. The B minima extensions both exhibit tightly coiled helices. The tightly coiled nature of these oligosaccharides may occlude residues and phosphate groups therefore making antibody access difficult.

These conformations demonstrate that, given similarly arranged phosphate groups, the differences in global minimum energy conformations between the G12R and G13R are unlikely to impact greatly on the conformation of 19A and 19F polysaccharide repeating unit chains. The B minima however produce structures that could perceivably cause difficulty with respect to vaccine efficacy. The reasons for the lack of cross protection observed between 19A and 19F in vaccines such as PCV7 are not readily apparent.

Chapter 7: Conclusions and Future Work

This work stands as both a comparison of the CHARMM and GLYCAM carbohydrate force fields and an initial computational investigation into the *S. pneumoniae* serotype 19A and 19F CPSs.

While there were discrepancies between the CHARMM and GLYCAM conformations, overall their predictions were in agreement. Both force fields presented minimum energy conformations in close agreement with one another even when extended to form oligosaccharides. GLYCAM consistently described a more flexible glycosidic linkage than CHARMM, with broader secondary minima of lower energy. CHARMM also predicted a far more readily accessible, albeit narrower, B region than GLYCAM. While the exact causes are difficult to isolate, these discrepancies could be explained by GLYCAM's lack of aliphatic hydrogen charges which are present in CHARMM.

The G12R and G13R linkages differed considerably. The G12R linkage was predicted by both force fields to have greater flexibility than that observed in G13R. Minimum energy conformations and the ability to transition between them also differed significantly. The difference was most pronounced under CHARMM. Both G12R and G13R global minima oligosaccharide extensions exhibited flat extended helical structures. The slight structural variations observed in the oligosaccharide extensions do not explain the lack of cross-protection observed for 19A by 19F vaccine conjugates.

This study confirms through both the CHARMM and GLYCAM force fields, that there are differences between the minimum energy conformations of the G12R and G13R disaccharides. While these differences could account for the lack of 19F vaccine conjugate cross protection for 19A, more work will need to be performed to confirm this. Future work simulating full 19A and 19F CPSs in solution using parameter sets including phosphodiester bond parameters allowing for the consideration of more extensive interresidue interactions would be beneficial. The addition of other dominant carbohydrate force fields in addition to CHARMM and GLYCAM (such as GROMOS), along with

experimental approaches such as NMR spectroscopy investigations into CPS structures could serve to validate further research. This investigation serves as an initial step into a multifaceted topic that could easily form the basis for a larger and more comprehensive study.

Chapter 8: References

1. Kuttel M. The Conformational Free Energy of Carbohydrates. *Mini Rev Org Chem*. 2011;8(3):256-262. doi:10.2174/157019311796197364.
2. Hansen H, Hünenberger P. A Reoptimized GROMOS Force Field for Hexopyranose-Based Carbohydrates Accounting for the Relative Free Energies of Ring Conformers, Anomers, Epimers, Hydroxymethyl Rotamers, and Glycosidic Linkage Conformers. *J Comput Chem*. 2011;32(6):998-1032. doi:10.1002/jcc.
3. Fadda E, Woods RJ. Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov Today*. 2010;15(15-16):596-609. doi:10.1016/j.drudis.2010.06.001.
4. Ciuffreda P, Colombo D, Ronchetti F, Toma L. Conformation analysis of the trisaccharide components of the repeating units of the capsular polysaccharides of *Streptococcus pneumoniae* types 19F and 19A. *Carbohydr Res*. 1992;232:327-339.
5. Brady JW, Schmidt RK. The role of hydrogen bonding in carbohydrates: molecular dynamics simulations of maltose in aqueous solution. *J Phys Chem*. 1993;97(4):958-966. doi:10.1021/j100106a024.
6. Pereira CS, Kony D, Baron R, Müller M, van Gunsteren WF, Hünenberger PH. Conformational and dynamical properties of disaccharides in water: a molecular dynamics study. *Biophys J*. 2006;90(12):4337-44. doi:10.1529/biophysj.106.081539.
7. Ponder JW, Case DA. Force fields for protein simulations. *Adv Protein Chem*. 2003;66:27-85. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2997720&tool=pmcentrez&rendertype=abstract>.
8. Scott WRP, Hünenberger PH, Tironi IG, et al. The GROMOS Biomolecular Simulation Program Package. *J Phys Chem A*. 1999;103(19):3596-3607.
9. Brooks BR, Iii CLB, Mackerell ADJ, et al. CHARMM : The Biomolecular Simulation Program. *J Comput Chem*. 2009;30(10):1545-1614. doi:10.1002/jcc.
10. Kirschner KN, Yongye AB, Tschampel SM, et al. GLYCAM06 : A Generalizable Biomolecular Force Field . *J Comput Chem*. 2008;29:622-655. doi:10.1002/jcc.
11. Ha SN, Giammona A, Field M, Brady JW. A revised potential-energy surface for molecular mechanics studies of carbohydrates. *Carbohydr Res*. 1988;180(2):207-221.
12. Woods RJ, Dwek RA, Edge CJ, Fraser-reid B. Molecular Mechanical and Molecular Dynamical Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM-93 Parameter Development. *J Phys Chem*. 1995;99(11):3832-3846.
13. Stortz CA, Johnson GP, French AD, Csonka GI. Comparison of different force fields for the study of disaccharides. *Carbohydr Res*. 2009;344(16):2217-28. doi:10.1016/j.carres.2009.08.019.
14. Oostenbrink C, Soares TA, van der Vegt NFA, van Gunsteren WF. Validation of the 53A6 GROMOS force field. *Eur Biophys J*. 2005;34(4):273-84. doi:10.1007/s00249-004-0448-6.
15. Lins RD, Hünenberger PH. A new GROMOS force field for hexopyranose-based carbohydrates.

- J Comput Chem.* 2005;26(13):1400-12. doi:10.1002/jcc.20275.
16. Guvench O, Greene SN, Kamath G, et al. Additive Empirical Force Field for Hexopyranose Monosaccharides. *J Comput Chem.* 2008;29(15):2543. doi:10.1002/jcc.
 17. Guvench O, Hatcher E, Venable R, Pastor R, Mackerell A. CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J Chem Theory Comput.* 2009;5(9):2353-2370. doi:10.1021/ct900242e.CHARMM.
 18. Alderson MR, Maisonneuve J, Newhouse LC, Boslego JW. PATH pneumococcal vaccine project. *Ped Health.* 2010;4(5):471-478. doi:10.2217/phe.10.53.
 19. Bogaert D, Hermans PWM, Adrian P V, Rümke HC, de Groot R. Pneumococcal vaccines: an update on current strategies. *Vaccine.* 2004;22(17-18):2209-20. doi:10.1016/j.vaccine.2003.11.038.
 20. Kuttel M, Brady JW, Naidoo KJ. Carbohydrate solution simulations: producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. *J Comput Chem.* 2002;23(13):1236-43. doi:10.1002/jcc.10119.
 21. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol.* 1963;7(1):95-99.
 22. Widmalm G. A perspective on the primary and three-dimensional structures of carbohydrates. *Carbohydr Res.* 2013;378:123-32. doi:10.1016/j.carres.2013.02.005.
 23. Kuttel M, Ravenscroft N, Foschiatti M, Cescutti P, Rizzo R. Conformational properties of two exopolysaccharides produced by *Inquilinus limosus*, a cystic fibrosis lung pathogen. *Carbohydr Res.* 2012;350:40-8. doi:10.1016/j.carres.2011.12.025.
 24. Laio A, Parrinello M. Escaping free-energy minima. *Proc Natl Acad Sci U S A.* 2002;99(20):12562-6. doi:10.1073/pnas.202427399.
 25. Jin P, Kong F, Xiao M, et al. First report of putative *Streptococcus pneumoniae* serotype 6D among nasopharyngeal isolates from Fijian children. *J Infect Dis.* 2009;200(9):1375-80. doi:10.1086/606118.
 26. Hausdorff WP, Dagan R, Beckers F, Schuerman L. Estimating the direct impact of new conjugate vaccines against invasive pneumococcal disease. *Vaccine.* 2009;27(52):7257-69. doi:10.1016/j.vaccine.2009.09.111.
 27. Cherian T. WHO expert consultation on serotype composition of pneumococcal conjugate vaccines for use in resource-poor developing countries, 26-27 October 2006, Geneva. *Vaccine.* 2007;25(36):6557-64. doi:10.1016/j.vaccine.2007.06.044.
 28. Schranz J. Pneumococcal conjugate vaccines: what do we know and what do we need? *Procedia Vaccinol.* 2009;1(1):189-205. doi:10.1016/j.provac.2009.07.032.
 29. Elberse K, Witteveen S, van der Heide H, et al. Sequence diversity within the capsular genes of *Streptococcus pneumoniae* serogroup 6 and 19. *PLoS One.* 2011;6(9):e25018. doi:10.1371/journal.pone.0025018.
 30. Jennings HJ, Rosell K-G, Carlo DJ. Structural determination of the capsular polysaccharide of *Streptococcus pneumoniae* type-19 (19F). *Can J Chem.* 1980;58(11):1069-1074.
 31. Katzenellenbogen E, Jennings HJ. Structural determination of the capsular polysaccharide of *Streptococcus pneumoniae* type 19A (57). *Carbohydr Res.* 1983;124(2):235-245.

32. Jakobsen H, Sigurdsson VD, Sigurdardottir S, Schulz D, Jonsdottir I. Pneumococcal Serotype 19F Conjugate Vaccine Induces Cross-Protective Immunity to Serotype 19A in a Murine Pneumococcal Pneumonia Model. *Infect Immun*. 2003;71(5):2956-2959. doi:10.1128/IAI.71.5.2956.
33. Hausdorff WP, Hoet B, Schuerman L. Do pneumococcal conjugate vaccines provide any cross-protection against serotype 19A? *BMC Pediatr*. 2010;10:4. doi:10.1186/1471-2431-10-4.
34. Aguiar SI, Brito MJ, Gonçalo-Marques J, Melo-Cristino J, Ramirez M. Serotypes 1, 7F and 19A became the leading causes of pediatric invasive pneumococcal infections in Portugal after 7 years of heptavalent conjugate vaccine use. *Vaccine*. 2010;28(32):5167-73. doi:10.1016/j.vaccine.2010.06.008.
35. Hanage WP. Serotype replacement in invasive pneumococcal disease: where do we go from here? *J Infect Dis*. 2007;196(9):1282-4. doi:10.1086/521630.
36. Pletz MW, Maus U, Krug N, Welte T, Lode H. Pneumococcal vaccines: mechanism of action, impact on epidemiology and adaption of the species. *Int J Antimicrob Agents*. 2008;32(3):199-206. doi:10.1016/j.ijantimicag.2008.01.021.
37. McIntosh EDG, Reinert RR. Global prevailing and emerging pediatric pneumococcal serotypes. *Expert Rev Vaccines*. 2011;10(1):109-29. doi:10.1586/erv.10.145.
38. World Health Organisation. *WHO Position Paper on Pneumococcal Vaccines 2012 Summary*; 2012:1-2.
39. Collier G, Vellore NA, Yancey JA, Stuart SJ, Latour RA. Comparison between empirical protein force fields for the simulation of the adsorption behavior of structured LK peptides on functionalized surfaces. *Biointerphases*. 2012;7(1-4):24. doi:10.1007/s13758-012-0024-z.
40. Guvench O, MacKerell Jr AD. Comparison of protein force fields for molecular dynamics simulations. *Methods Mol Biol*. 2008;443:63-88. doi:10.1007/978-1-59745-177-2_4.
41. Hu H, Elstner M, Hermans J. Comparison of a QM/MM Force Field and Molecular Mechanics Force Fields in Simulations of Alanine and Glycine “Dipeptides” (Ace-Ala-Nme and Ace-Gly-Nme) in Water in Relation to the Problem of Modeling the Unfolded Peptide Backbone in Solution. *PROTEINS Struct Funct Genet*. 2003;50(3):451-463.
42. Perez S, Imberty A, Engelsen SB, et al. A comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates. *Carbohydr Res*. 1998;314:141-155.
43. Hemmingsen L, Madsen DE, Esbensen AL, Olsen L, Engelsen SB. Evaluation of carbohydrate molecular mechanical force fields by quantum mechanical calculations. *Carbohydr Res*. 2004;339(5):937-48. doi:10.1016/j.carres.2003.11.024.
44. Bonaccorsi F, Catelani G, Oscarson S. A new route for the synthesis of Streptococcus pneumoniae 19F and 19A capsular polysaccharide fragments avoiding the beta-mannosamine glycosylation step. *Carbohydr Res*. 2009;344(12):1442-8. doi:10.1016/j.carres.2009.04.012.
45. Legnani L, Ronchi S, Fallarini S, et al. Synthesis, molecular dynamics simulations, and biology of a carba-analogue of the trisaccharide repeating unit of Streptococcus pneumoniae 19F capsular polysaccharide. *Org Biomol Chem*. 2009;7(21):4428-36. doi:10.1039/b911323a.
46. Goldblatt D. Conjugate Vaccines. *Clin Exp Immunol*. 2000;119(1):1-3.
47. Mathews CJ, van Holde KE, Ahern KG. *Biochemistry*. 3rd ed. San Francisco: Benjamin-

Cummings; 2000.

48. Harding MM, Anderberg PI, Haymet ADJ. "Antifreeze" glycoproteins from polar fish. *Eur J Biochem.* 2003;270(7):1381-1392. doi:10.1046/j.1432-1033.2003.03488.x.
49. Varki a. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology.* 1993;3(2):97-130. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8490246>.
50. Autieri E, Sega M, Pederiva F. Puckering Free Energy of Pyranoses Using a Combined Metadynamics–Umbrella Sampling Approach. *J Chem Phys.* 2010;133(9).
51. Weintraub A. Immunology of bacterial polysaccharide antigens. *Carbohydr Res.* 2003;338(23):2539-2547. doi:10.1016/j.carres.2003.07.008.
52. Tamaru Y, Takani Y, Yoshida T, Sakamoto T. Crucial Role of Extracellular Polysaccharides in Desiccation and Freezing Tolerance in the Terrestrial Cyanobacterium *Nostoc commune* Crucial Role of Extracellular Polysaccharides in Desiccation and Freezing Tolerance in the Terrestrial Cyanobacterium *Nost.* *Appl Environ Microbiol.* 2005;71(11):7327-7333. doi:10.1128/AEM.71.11.7327.
53. Ophir T, Gutnick DL. A role for exopolysaccharides in the protection of microorganisms from desiccation. *Appl Environ Microbiol.* 1994;60(2):740-5. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=201377&tool=pmcentrez&rendertype=abstract>.
54. Avci FY, Kasper DL. How bacterial carbohydrates influence the adaptive immune system. *Annu Rev Immunol.* 2010;28:107-30. doi:10.1146/annurev-immunol-030409-101159.
55. Finland M, Dowling HF. Cutaneous Reactions and Antibody Response to Intracutaneous Injections of *Pneumococcus* Polysaccharides. *J Immunol.* 1935;29:285-299.
56. MacLeod CM, Hodges RG, Heidelberger M, Bernhard WG. Prevention of Pneumococcal Pneumoniae by Immunization with Specific Capsular Polysaccharides. *J Exp Med.* 1945;82(6):445.
57. Goebel WF, Avery OT. Chemo-Immunological Studies on Conjugated Carbohydrate-Proteins. *J Exp Med.* 1929;50(4):521-531.
58. Grabenstein JD, Klugman KP. A century of pneumococcal vaccination research in humans. *Clin Microbiol Infect.* 2012;18 Suppl 5:15-24. doi:10.1111/j.1469-0691.2012.03943.x.
59. Austrian R. A brief history of pneumococcal vaccines. *Drugs Aging.* 1999;15 Suppl 1:1-10. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10690790>.
60. AlonsoDeVelasco E, Verheul a F, Verhoef J, Snippe H. *Streptococcus pneumoniae*: virulence factors, pathogenesis, and vaccines. *Microbiol Rev.* 1995;59(4):591-603. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=239389&tool=pmcentrez&rendertype=abstract>.
61. Butler JC, Shapiro ED, Carlone GM. Pneumococcal vaccines: history, current status, and future directions. *Am J Med.* 1999;107(1A):69S-76S. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10451012>.
62. Kazanjian P. Changing Interest among Physicians toward Pneumococcal Vaccination throughout the Twentieth Century. *J Hist Med Allied Sci.* 2004;59(4):555-587.
63. Merck. PNEUMOVAX23 (Pneumococcal Vaccine Polyvalent). 2013. Available at:

<https://www.merckvaccines.com/Products/Pneumovax/Pages/home>. Accessed January 20, 2013.

64. Johnson HL, Deloria-Knoll M, Levine OS, et al. Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: the pneumococcal global serotype project. *PLoS Med.* 2010;7(10). doi:10.1371/journal.pmed.1000348.
65. Skinner JM, Indrawati L, Cannon J, et al. Pre-clinical evaluation of a 15-valent pneumococcal conjugate vaccine (PCV15-CRM197) in an infant-rhesus monkey immunogenicity model. *Vaccine.* 2011;29(48):8870-6. doi:10.1016/j.vaccine.2011.09.078.
66. Mallajosyula SS, Guvench O, Hatcher E, Mackerell AD. CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates. *J Chem Theory Comput.* 2012;8(2):759-776. doi:10.1021/ct200792v.
67. Grootenhuis PDJ, Haasnoot CAG. A CHARMM Based Force Field for Carbohydrates Using the CHEAT Approach: Carbohydrate Hydroxyl Groups Represented by Extended Atoms. *Mol Simul.* 1993;10(2-6):75-95.
68. Pérez S, Kouwijzer M, Mazeau K, Engelsens SB. Modeling polysaccharides: present status and challenges. *J Mol Graph.* 1996;14(6):307-21, 361-2. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9195482>.
69. Reiling S, Schlenkrich M, Brickmann J. Force Field Parameters for Carbohydrates. *J Comput Chem.* 1996;17(4):450-468.
70. Raman EP, Guvench O, MacKerell AD. CHARMM additive all-atom force field for glycosidic linkages in carbohydrates involving furanoses. *J Phys Chem B.* 2010;114(40):12981-94. doi:10.1021/jp105758h.
71. Woods Group Complex Carbohydrate Research Center University of Georgia. GLYCAM Web. 2005. Available at: <http://www.glycam.com>. Accessed February 3, 2011.
72. Glennon TM, Zheng Y-J, Grand SM Le, Shutzberg BA, Merz KMJ. A force field for monosaccharides and (1 → 4) linked polysaccharides. *J Comput Chem.* 1994;15(9):1019-1040.
73. Homans SW. A molecular mechanical force field for the conformational analysis of oligosaccharides: comparison of theoretical and crystal structures of Man alpha 1-3Man beta 1-4GlcNAc. *Biochemistry.* 1990;29(39):9110-8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2271581>.
74. Van Gunsteren WF, Brenendsen HJC. Groningen Molecular Simulation (GROMOS) Library Manual. 1987:1-221.
75. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem.* 2004;25(13):1656-76. doi:10.1002/jcc.20090.
76. Kräutler V, Müller M, Hünenberger PH. Conformation, dynamics, solvation and relative stabilities of selected beta-hexopyranoses in water: a molecular dynamics study with the GROMOS 45A4 force field. *Carbohydr Res.* 2007;342(14):2097-124.
77. Christen M, Kunz A-PE, van Gunsteren WF. Sampling of rare events using hidden restraints. *J Phys Chem B.* 2006;110(16):8488-98. doi:10.1021/jp0604948.
78. Perić L, Pereira CS, Pérez S, Hünenberger PH. Conformation, dynamics and ion-binding

- properties of single-chain polyuronates: a molecular dynamics study. *Mol Simul.* 2008;34(4):421-446. doi:10.1080/08927020701759699.
79. Oliveira O V, Freitas LCG, Straatsma TP, Lins RD. Interaction between the CBM of Cel9A from *Thermobifida fusca* and cellulose fibers. *J Mol Recognit.* 2009;22(1):38-45. doi:10.1002/jmr.925.
 80. Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comput Phys.* 1977;23(2):187.
 81. Darve E, Rodríguez-Gómez D, Pohorille A. Adaptive biasing force method for scalar and vector free energy calculations. *J Chem Phys.* 2008;128(14):144120. doi:10.1063/1.2829861.
 82. Laio A, Gervasio FL. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports Prog Phys.* 2008;71(12):126601. doi:10.1088/0034-4885/71/12/126601.
 83. Kuttel M, Naidoo KJ. Free energy surfaces for the alpha(1 --> 4)-glycosidic linkage: implications for polysaccharide solution structure and dynamics. *J Phys Chem B.* 2005;109(15):7468-74. doi:10.1021/jp044756m.
 84. Spiwok V, Tvaroska I. Metadynamics modelling of the solvent effect on primary hydroxyl rotamer equilibria in hexopyranosides. *Carbohydr Res.* 2009;344(12):1575-81. doi:10.1016/j.carres.2009.05.019.
 85. Biarne X, Ardevol A, Planas A, Rovira C, Laio A, Parrinello M. The Conformational Free Energy Landscape of beta-D-Glucopyranose. Implications for Substrate Preactivation in beta-Glucoside Hydrolases. *J Am Chem Soc.* 2007;129(35):10686-10693.
 86. Spiwok V, Králová B, Tvaroska I. Modelling of beta-D-glucopyranose ring distortion in different force fields: a metadynamics study. *Carbohydr Res.* 2010;345(4):530-7. doi:10.1016/j.carres.2009.12.011.
 87. Spiwok V, D. French A. Modelling the Effect of Solvents on Carbohydrates. *Mini Rev Org Chem.* 2011;8(3):249-255. doi:10.2174/157019311796197445.
 88. Bonomi M, Parrinello M. Enhanced Sampling in the Well-Tempered Ensemble. 2010;104(19):190601.
 89. Donadio D, Raiteri P, Parrinello M. Topological Defects and Bulk Meltin of Hexagonal Ice. 2005;109(12):5421-5424.
 90. Micheletti C, Laio A, Parrinello M. Reconstructing the Density of States by History-Dependent Metadynamics. *Phys Rev Lett.* 2004;92(17):170601. doi:10.1103/PhysRevLett.92.170601.
 91. Mulholland K. Strategies for the control of pneumococcal diseases. *Vaccine.* 1999;17 Suppl 1:S79-84. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10471187>.
 92. Prymula R. Re: Global serotype distribution among *Streptococcus pneumoniae* isolates causing otitis media in children: potential implications for pneumococcal conjugate vaccines. *Vaccine.* 2009;27(35):4739-40. doi:10.1016/j.vaccine.2009.05.098.
 93. Jiang S, Wang LEI, Reeves PR. Molecular Characterization of *Streptococcus pneumoniae* Type 4 , 6B, 8 and 8C Capsular Polysaccharide Gene Clusters. *Infect Immun.* 2001;69(3):1244-1255. doi:10.1128/IAI.69.3.1244.
 94. Hadler J, Reingold A, Thomas AR, et al. Effect of Introduction of the Pneumococcal Conjugate

Vaccine on Drug-Resistant *Streptococcus pneumoniae*. *N Engl J Med*. 2006;354(14).

95. Reinert R, Jacobs MR, Kaplan SL. Pneumococcal disease caused by serotype 19A: review of the literature and implications for future vaccine development. *Vaccine*. 2010;28(26):4249-59. doi:10.1016/j.vaccine.2010.04.020.
96. Morona JK, Morona R, Paton JC. Comparative genetics of capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* types belonging to serogroup 19. *J Bacteriol*. 1999;181(17):5355-64. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=94042&tool=pmcentrez&rendertype=abstract>.
97. Reingold A, Farley MM, Harrison L, et al. Invasive Pneumococcal Disease in Children 5 Years After Conjugate Vaccine Introduction---Eight States, 1998--2005. *Morb Mortal Wkly Rep*. 2008;57(6):144-8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18272956>.
98. Van Gils EJM, Veenhoven RH, Hak E, et al. Pneumococcal Conjugate Vaccination and Nasopharyngeal Acquisition of Pneumococcal Serotype 19A Strains. *J Am Med Assoc*. 2010;304(10):1099-1106.
99. Kaplan SL, Barson WJ, Lin PL, et al. Serotype 19A Is the most common serotype causing invasive pneumococcal infections in children. *Pediatrics*. 2010;125(3):429-36. doi:10.1542/peds.2008-1702.
100. Richter SS, Heilmann KP, Dohrn CL, Riahi F, Diekema DJ, Doern G V. Pneumococcal Serotypes before and after Introduction of Conjugate Vaccines. *Emerg Infect Dis*. 2013;19(7):1074-1083.
101. By A, Sobocki P, Forsgren A, Silfverdal S-A. Comparing health outcomes and costs of general vaccination with pneumococcal conjugate vaccines in Sweden: a Markov model. *Clin Ther*. 2012;34(1):177-89. doi:10.1016/j.clinthera.2011.12.007.
102. Strutton DR, Farkouh R a, Earnshaw SR, et al. Cost-effectiveness of 13-valent pneumococcal conjugate vaccine: Germany, Greece, and The Netherlands. *J Infect*. 2012;64(1):54-67. doi:10.1016/j.jinf.2011.10.015.
103. Bock K. The Preferred Conformation of Oligosaccharides in Solution Inferred from High Resolution NMR Data and Hard Sphere Exo-Anomeric Calculation. *Pure Appl Chem*. 1983;55(4):605-622.
104. Tai JC, Allinger NL. Molecular Mechanics Calculations on Conjugated Nitrogen-Containing Heterocycles. *J Am Chem Soc*. 1988;110(8):2050-2055.
105. Hocquet A, Langgård M. An Evaluation of the MM+ Force Field. *J Mol Model*. 1998;4(3):94-112. doi:10.1007/s008940050128.
106. Klok RM, Lindkvist R-M, Ekelund M, Farkouh RA, Strutton DR. Cost-effectiveness of a 10- versus 13-valent pneumococcal conjugate vaccine in Denmark and Sweden. *Clin Ther*. 2013;35(2):119-34. doi:10.1016/j.clinthera.2012.12.006.
107. Lee CJ, Fraser B a, Szu S, Lin KT. Chemical structure of and immune response to polysaccharides of *Streptococcus pneumoniae*. *Rev Infect Dis*. 1981;3(2):323-31. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7020047>.
108. Grant LR, O'Brien SE, Burbidge P, et al. Comparative immunogenicity of 7 and 13-valent pneumococcal conjugate vaccines and the development of functional antibodies to cross-reactive serotypes. *PLoS One*. 2013;8(9):e74906. doi:10.1371/journal.pone.0074906.

109. Bryant KA, Block SL, Baker SA, Gruber WC, Scott DA. Safety and immunogenicity of a 13-valent pneumococcal conjugate vaccine. *Pediatrics*. 2010;125(5):866-75. doi:10.1542/peds.2009-1405.
110. Jefferies JMC, Macdonald E, Faust SN, Clarke SC. 13-valent pneumococcal conjugate vaccine (PCV13). *Hum Vaccin*. 2011;7(10):1012-8. doi:10.4161/hv.7.10.16794.
111. Poolman J, Frasch C, Nurkka A, Käyhty H, Biemans R, Schuerman L. Impact of the conjugation method on the immunogenicity of *Streptococcus pneumoniae* serotype 19F polysaccharide in conjugate vaccines. *Clin Vaccine Immunol*. 2011;18(2):327-36. doi:10.1128/CVI.00402-10.
112. Smith KJ, Wateska AR, Nowalk MP, Raymund M, Nuorti JP, Zimmerman RK. Cost-effectiveness of adult vaccination strategies using pneumococcal conjugate vaccine compared with pneumococcal polysaccharide vaccine. *JAMA*. 2012;307(8):804-12. doi:10.1001/jama.2012.169.
113. Nelson MT, Humphrey W, Gursoy a., et al. NAMD: a Parallel, Object-Oriented Molecular Dynamics Program. *Int J High Perform Comput Appl*. 1996;10(4):251-268. doi:10.1177/109434209601000401.
114. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26(16):1781-802. doi:10.1002/jcc.20289.
115. Gordon M. StrAP. 2012. Available at: <https://github.com/marcgrdn/StrAP>.
116. Gordon M. extractFrames. 2013. Available at: <https://github.com/marcgrdn/extractFrames>.
117. Gullingsrud J. CatDCD - Concatenate DCD files. *Theor Comput Biophys Group, Univ Illinois Urbana-Champaign*. 2009. Available at: <http://www.ks.uiuc.edu/Development/MDTools/catdcd/>.
118. Kuttel MM. Conformational free energy maps for globobiose (alpha-D-Galp-(1-->4)-beta-D-Galp) in implicit and explicit aqueous solution. *Carbohydr Res*. 2008;343(6):1091-8. doi:10.1016/j.carres.2008.01.038.
119. Yadav M. *Computational Biochemistry*. New Delhi: Discovery Publishing House; 2009:45.

Chapter 9: Appendices

Appendix A: NAMD Configuration Files

1. *G12R_C Unbiased Simulation Configuration File*

```
# Written by: M. Kuttel November 2011
# Modified by: M. Gordon February 2012
# NAMD Config file for unbiased run of aDGlc-a12-aLRha bond of S. pneumoniae serogroup 19 carbohydrate repeating unit

# input
coordinates      ./structures/aDGlc-a12-aLRha.pdb
structure        ./structures/aDGlc-a12-aLRha.psf

#used for restarting simulation using minimisation output files
bincoordinates ./minimisation/output/aDGlc-a12-aLRha.coor
binvelocities  ./minimisation/output/aDGlc-a12-aLRha.vel
extendedSystem ./minimisation/output/aDGlc-a12-aLRha.xsc

parameters      ./forcefields/toppar_carb_apr12/par_all36_carb(marcEdit).prm
paratypecharm   on

# output
set output      ./output/aDGlc-a12-aLRha
outputname      $output
DCDfile         $output.dcd
#by default it should be binary output
binaryoutput    yes
outputEnergies  100
dcdfreq        1000

#fixedAtoms     off

# Basic dynamics
exclude         scaled1-4
1-4scaling      1
COMmotion       no
dielectric      1.0

# Simulation space partitioning
switching       on
switchdist      12
cutoff          15
```

pairlistdist 18

```
# Temperature control
reassignFreq 1000
reassignTemp 25
reassignIncr 25
reassignHold 300
```

```
# run duration
run 100000000
```

2. G12R_C Biased Simulation Configuration File

Written by: M. Kuttel November 2011

Modified by: M. Gordon February 2012

NAMD Config file for Metadynamics run of aDGlc-a12-aLRha bond of S. pneumoniae serogroup 19 carbohydrate repeating unit

input

```
coordinates ./structures/aDGlc-a12-aLRha.pdb
structure ./structures/aDGlc-a12-aLRha.psf
```

#used for restarting simulation using minimisation output files

bincoordinates ./minimisation/output/aDGlc-a12-aLRha.coor

binvelocities ./minimisation/output/aDGlc-a12-aLRha.vel

extendedSystem ./minimisation/output/aDGlc-a12-aLRha.xsc

parameters ./forcefields/toppar_carb_apr12/par_all36_carb(marcEdit).prm

paratypecharmm on

output

set output ./output/aDGlc-a12-aLRha

outputname \$output

DCDfile \$output.dcd

binaryoutput yes

outputEnergies 100

dcdfreq 1000

#fixedAtoms off

Basic dynamics

exclude scaled1-4

1-4scaling 1

```

COMmotion      no
dielectric      1.0

# Simulation space partitioning
switching      on
switchdist     12
cutoff         15
pairlistdist    18

# Temperature control
reassignFreq 1000
reassignTemp 25
reassignIncr 25
reassignHold 300

# NAMD colvars module
colvars on
colvarsConfig colvars.txt

# run duration
run 1500000000

```

Appendix B: NAMD Colvars Module Configuration Files

1. *G12R_C*

```

colvarsTrajFrequency 1000
#NB - change this when change targetNumSteps
colvarsTrajAppend off
#when running consecutive simulations with the same outputName
#enable this option to preserve the previous contents of the trajectory file.

colvar {
  name Phi
  width 2.5
  dihedral {
    group1 {
      atomnumbers { 24 }
    }
    group2 {
      atomnumbers { 23 }
    }
    group3 {
      atomnumbers { 2 }
    }
  }
}

```

```

    group4 {
        atomnumbers { 1 }
    }
}
lowerBoundary -180
upperBoundary 180
}

```

```

colvar {
    name Psi
    width 2.5
    dihedral {
        group1 {
            atomnumbers { 10 }
        }
        group2 {
            atomnumbers { 1 }
        }
        group3 {
            atomnumbers { 2 }
        }
        group4 {
            atomnumbers { 23 }
        }
    }
    lowerBoundary -180
    upperBoundary 180
}

```

```

Metadynamics {
    name metaPhiPsi
    colvars Phi Psi
    hillWeight 0.05 #choose carefully - 0.01 seemed too small, 0.5 seemed to cause instability
    dumpFreeEnergyFile yes
}

```